

DEA de Sciences Cognitives
Écoles des Hautes Études en Sciences Sociales
1996-1997

**Reconnaissance d'objets dans des scènes
naturelles**

**Rôle des informations sensorielles et
contextuelles**

Guillaume Giraudet

Le 11 septembre 1997

Travaux réalisés sous la direction de Corinne Roumes et Jean-Pierre Menu

Rapporteur : Kevin O'Regan

TABLE DES MATIÈRES

1- INTRODUCTION	1
2- POSITION DU PROBLÈME	2
3- PROTOCOLE EXPÉRIMENTAL	9
3-1- MATERIEL	9
3-2- STIMULATIONS.....	10
3-3- VARIABLES EXPERIMENTALES	11
3-3-1- Variation du contenu fréquentiel	11
3-3-2- Variation du contenu chromatique.....	12
3-3-3- Variation du contenu contextuel.....	12
3-3-4- Contrôle du temps de présentation	13
3-3-5- Synthèse et notation des différentes variables	13
3-4- DEROULEMENT TEMPOREL.....	14
3-4-1- Organisation générale de l'expérimentation.....	14
3-4-2- Déroulement d'une présentation	14
3-5- SUJETS.....	15
3-6- PARAMETRES RECUEILLIS.....	15
3-7- ANALYSE DES RESULTATS	15
4- RÉSULTATS	16
4-1- TAUX D'ERREURS	16
4-1-1- Taux d'erreurs sur les images d'origine.....	16
4-1-2- Variabilité liée aux sujets et aux images	16
4-1-3- Facteurs de variation sur l'ensemble des données.....	18
4-1-3-1- Effet de la dégradation de l'image	18
4-1-3-2- Effet du temps de présentation	20
4-1-3-3- Effet du contexte	21
4-1-4- Variation entre la 1 ^{ère} et la 3 ^{ème} répétition.....	24
4-1-4-1- Taux d'erreurs moyen	24
4-1-4-2- Effet d'une altération dans la prise d'information visuelle	24
4-1-4-3- Effet du contexte	26
4-1-5- Cas particulier de certaines images.....	26
4-2- TEMPS DE REPONSE.....	27
4-2-1- Variabilité liée aux sujets et aux images	28
4-2-2- Facteurs de variation sur l'ensemble des données.....	29
4-2-2-1- Effet de la dégradation de l'image	29
4-2-2-2- Effet du temps de présentation	30
4-2-2-3- Effet du contexte	30
4-2-3- Variation entre la 1 ^{ère} et la 3 ^{ème} répétition.....	32
4-2-3-1- Temps de réponse moyen	32
4-2-3-2- Effet d'une altération dans la prise d'information visuelle	33
4-2-3-3- Effet du contexte	34
5- DISCUSSION	34
5-1- METHODOLOGIE	35
5-2- DEGRADATION DU CONTENU PHYSIQUE DE LA STIMULATION ET CONTEXTE.....	35
5-3- APPRENTISSAGE ET FAMILIARITE.....	38

6- CONCLUSIONS ET PERSPECTIVES	42
6-1- PERSPECTIVES A COURT TERME	43
6-2- PERSPECTIVES A LONG TERME	44
7- BIBLIOGRAPHIE	45

LISTE DES FIGURES

<u>Figure 1</u> : schéma général d'organisation du système visuel	6
<u>Figure 2</u> : points d'impact des facteurs de variation	9
<u>Figure 3</u> : Méthode de sélection des images présentées	10
<u>Figure 4</u> : Contenu fréquentiel résultant des filtrages en fréquences spatiales.....	12
<u>Figure 5</u> : Modification du contenu contextuel	13
<u>Figure 6</u> : Déroulement d'une séquence.....	15
<u>Figure 7</u> : taux d'erreurs moyen des sujets.....	17
<u>Figure 8</u> : Taux d'erreurs moyen sur les différentes images.	17
<u>Figure 9</u> : Effet du filtrage en fréquences spatiales.	18
<u>Figure 10</u> : Effet du filtrage fréquentiel sur le taux d'erreurs moyen obtenu avec les différentes images.....	19
<u>Figure 11</u> : Interdépendance des filtrages fréquentiels et chromatique sur le taux d'erreurs moyen.....	19
<u>Figure 12</u> : Effet du filtrage chromatique sur le taux d'erreurs pour les huit images.....	20
<u>Figure 13</u> : Interdépendance du filtrage fréquentiel et de la contrainte temporelle sur le taux d'erreurs moyen.....	20
<u>Figure 14</u> : Effet conjoint des restrictions dans la prise d'information	21
<u>Figure 15</u> : Effet du temps de présentation sur le taux d'erreurs moyen obtenu avec les différentes images.....	21
<u>Figure 16</u> : Effet du contexte sur le taux d'erreurs moyen.....	22
<u>Figure 17</u> : Effet du filtrage en fréquences spatiales sur le taux d'erreurs moyen	22
<u>Figure 18</u> : Effet du filtrage chromatique sur le taux d'erreurs moyen selon.....	23
<u>Figure 19</u> : Effet du temps de présentation sur le taux d'erreurs moyen selon	23
<u>Figure 20</u> : Effet du contexte sur le taux d'erreurs moyen pour les différentes images...23	
<u>Figure 21</u> : Variation du taux d'erreurs entre la 1ère et la 3ème série expérimentale.....24	
<u>Figure 22</u> : Effet du filtrage en fréquences spatiales sur le taux d'erreurs moyen	25
<u>Figure 23</u> : Effet du filtrage chromatique sur le taux d'erreurs moyen	25
<u>Figure 24</u> : Effet du temps de présentation sur le taux d'erreurs moyen.....	25
<u>Figure 25</u> : Effet du contexte sur le taux d'erreurs moyen à la 1ère et la 3ème série expérimentale.....	26
<u>Figure 26</u> : Temps de réponse moyen des sujets.	28
<u>Figure 27</u> : Temps de réponse pour les différentes images.....	28
<u>Figure 28</u> : Effet du filtrage en fréquences spatiales sur le temps de réponse.....	29
<u>Figure 29</u> : Effet du filtrage chromatique sur le temps de réponse.....	29
<u>Figure 30</u> : Effet du temps de présentation sur le temps de réponse.	30
<u>Figure 31</u> : Effet du contexte sur le temps de réponse.....	30
<u>Figure 32</u> : Effet du filtrage en fréquences spatiales sur le temps de réponse.....	31
<u>Figure 33</u> : Effet du filtrage chromatique sur le temps de réponse.....	31
<u>Figure 34</u> : Effet du temps de présentation sur le temps de réponse	32
<u>Figure 35</u> : Évolution du temps de réponse moyen au cours des séries expérimentales. 32	
<u>Figure 36</u> : Effet du filtrage en fréquences spatiales sur le temps de réponse.....	33
<u>Figure 37</u> : Effet du filtrage chromatique sur le temps de réponse	33
<u>Figure 38</u> : Effet du temps de présentation sur le temps de réponse	33
<u>Figure 39</u> : Effet du contexte sur le temps de réponse à la 1ère et la 3ème série expérimentale.....	34

LISTE DES TABLEAUX

<u>Tableau 1</u> : Variation de la luminance induite par le filtrage chromatique	12
<u>Tableau 2</u> : Taux de pointages en présence du distracteur.	27

1- INTRODUCTION

On constate quotidiennement que la vision est le sens qui nous fournit le plus d'informations sur le monde extérieur. En effet, c'est pour l'essentiel à partir de la perception visuelle des formes que l'on élabore une représentation de l'espace et que l'on détermine les interactions entre les différents éléments isolables d'une scène naturelle. La perception des formes est donc le meilleur moyen pour l'homme et en particulier pour le pilote d'appréhender son environnement. Cependant, dans les situations de vol de nuit ou de mauvaises conditions météorologiques, les moyens de suppléance proposés aux pilotes sont des systèmes d'aide qui permettent d'avoir des images mieux adaptées à la demande mais dont le contenu physique ne correspond plus aux habitudes perceptives de l'utilisateur. Ces dégradations qualitatives et quantitatives de l'information sensorielle diminuent les performances des pilotes nécessitant alors un entraînement réalisé sur des parcours reconnus de jour afin d'acquérir une familiarité de l'espace survolé. Il apparaît que les informations liées à l'environnement de la cible soient nécessaires à la bonne réalisation de l'opération. On sait en effet que les objets ne sont pas disposés aléatoirement dans la scène mais qu'il existe entre eux des relations spatiales cohérentes déterminées par leur nature. Les études ne portant que sur la perception de formes simples ou complexes isolées sur un fond homogène ne rendent donc pas compte des difficultés opérationnelles liées à la perception de formes complexes dans un environnement naturel.

Les multiples tâches (détection, reconnaissance ou identification) et stimulations (réseaux, camouflage, images ambiguës, formes géométriques, ...) qui peuvent être utilisées dans les expérimentations sur la perception des formes font appel à des niveaux de traitement très variés. Dans cette étude, le choix s'est porté sur des images de la vie quotidienne et une tâche de reconnaissance.

L'étude menée ici s'attache d'une part à contrôler les conditions d'altération des informations sensorielles, c'est-à-dire des informations fournies par l'environnement au système visuel à un instant donné, et d'autre part à déterminer les possibilités de maintien de la performance avec ou sans les informations liées à la connaissance du contexte, c'est-à-dire les informations sur l'arrangement spatial des éléments isolables d'une scène naturelle.

Un état de la littérature concernant d'une part les principales théories de la perception des formes et d'autre part les facteurs spécifiquement abordés dans l'étude permet de justifier le protocole expérimental développé ici. Les résultats recueillis sur une population de cinq sujets rendent compte de la performance de reconnaissance d'objets dans des scènes naturelles au fur et à mesure de la dégradation des informations sensorielles et contextuelles. Le rôle respectif des différents types d'information est précisé et analysé dans la discussion qui s'attache à mettre en évidence l'évolution de la stratégie des sujets au cours d'une exposition répétée à des images dégradées par rapport à celles qu'ils voient dans leur vie courante. Enfin, la dernière partie permet d'expliquer brièvement ce que l'on peut conclure de cette étude et quelles sont les perspectives envisageables pour en compléter les résultats.

2- POSITION DU PROBLÈME

Dans la plupart des modèles de la perception des formes, le schéma général d'organisation des processus de traitement de l'information visuelle est assez uniforme. Selon ce schéma, la perception s'organise selon deux voies de traitement : une voie ascendante, dite "bottom-up", qui achemine les données recueillies au niveau des capteurs oculaires, extrait les caractéristiques primitives du signal visuel et leur organisation spatio-temporelle ; une voie descendante, dite "top-down", qui provient des aires corticales de haut niveau pour aller préactiver de façon spécifique certains détecteurs de bas niveau et ainsi faciliter le traitement des données sensorielles.

Dans le cadre de ce modèle général, une expérimentation a été menée, chez l'homme, pour analyser la part respective des processus de traitement ascendants et descendants dans une tâche de reconnaissance d'objets dans des images naturelles.

Différentes tâches pour la perception des formes

Pour étudier la perception des formes, l'approche expérimentale considère différents types de tâches : la détection, la discrimination, la reconnaissance et l'identification de formes (Bonnet, 1986).

- Dans la détection, le sujet doit répondre à une question portant sur l'absence ou la présence d'un stimulus (seuil absolu).

- Dans la discrimination, le sujet doit répondre à une question portant sur l'absence ou la présence d'une différence entre deux stimulus (seuil différentiel).

- La reconnaissance est réservée à des paradigmes comportant deux phases successives dans lesquelles la tâche du sujet consiste à dire si chaque item ou chaque stimulus présenté au cours de la deuxième phase avait été, ou non, antérieurement présenté lors de la première phase.

- L'identification consiste à établir une correspondance bijective entre un ensemble de stimulus et un ensemble de réponses qui sont en général les étiquettes, les noms de ces stimulus.

La tâche de détection est un moyen privilégié d'étude de la voie ascendante, comme c'est le cas de la Fonction de Sensibilité aux Contrastes. Les modèles de la perception des formes basés sur une contribution relative des voies ascendante et descendante reposent essentiellement sur les autres types de tâches aux premiers rangs desquelles on trouve la recherche visuelle (tâche de discrimination) et l'identification de formes. Les divergences dans les divers modèles proposés portent sur le poids relatif des deux voies et leur relation, en fonction des différentes tâches considérées.

"Feature Integration Theory" (Treisman) et théories de la recherche visuelle

Treisman, dans une théorie générale de la perception des formes, "Feature Integration Theory" (FIT), distingue deux niveaux de traitement de l'information mis en évidence à la fois dans les tâches de recherche visuelle, où une cible définie doit être trouvée parmi un nombre variable de distracteurs, et dans les tâches d'identification (Treisman & Gelade, 1980 ; Treisman & Gormican, 1988).

Le premier niveau est dit "préattentif". Il consiste en une extraction en parallèle de l'ensemble des primitives visuelles de la stimulation, c'est un traitement très rapide

qui s'effectue de façon automatique. Le second niveau est qualifié d'"attentif". C'est un processus attentionnel qui, en se focalisant séquentiellement sur les différentes dimensions extraites, permet leur regroupement et leur localisation dans la scène pour établir une représentation structurée des objets et de leurs relations.

Le premier niveau est un flux informationnel ascendant (bottom-up) et le second un flux informationnel descendant (top-down). Ce type de théorie d'organisation du système visuel est utilisé non seulement pour analyser des données expérimentales (Treisman & Gelade, 1980 ; Treisman & Gormican, 1988) mais aussi comme base de fonctionnement de certains modèles neurobiologiques de reconnaissance de formes (Olshausen, Anderson & Van Essen, 1993).

Cependant, au niveau de la recherche visuelle, certains auteurs discutent l'indépendance des deux niveaux de traitement suggérée par la FIT. En effet, selon Treisman, si la tâche peut être effectuée à partir de la seule extraction des primitives, alors le second niveau n'intervient pas. La recherche est dite "efficace" puisqu'elle se fait en parallèle sur l'ensemble du champ de présentation et indépendamment du nombre de distracteurs. Selon Wolfe, par exemple, l'étape préattentive guide la seconde étape de focalisation de l'attention (Wolfe, Friedman-Hill & Bilsky 1994 ; Wolfe & Bennett, 1997). Le traitement en parallèle de la voie ascendante produit une carte d'activation pour chaque dimension indiquant ainsi à la seconde étape de traitement où se situe la cible. Les voies de traitement sont étroitement dépendantes.

Enfin, la "Similarity Theory" propose un modèle intermédiaire à ceux de Treisman et de Wolfe (Duncan & Humphreys, 1989). Selon ces auteurs, il existe un continuum entre les recherches parallèle et séquentielle, la recherche de la cible devenant de plus en plus difficile au fur et à mesure que la cible et les distracteurs deviennent similaires et que l'ensemble des distracteurs devient hétérogène.

Ces diverses théories établies sur une tâche de recherche visuelle ne permettent pas de rendre compte clairement du niveau d'intervention et des interrelations des voies ascendantes (automatiques, préattentives, parallèles) et descendantes (contrôlées, attentives, séquentielles).

Théories de la reconnaissance et de l'identification d'objets

Compte tenu des difficultés à transposer la FIT au niveau des données expérimentales sur la reconnaissance des formes et objets, Biederman propose un modèle qui décompose la perception en trois phases (Biederman, 1987) :

- la première phase consiste en une extraction des dimensions primitives de l'objet;
- ces dimensions activent ensuite des représentations de parties appelées "Géons" ;
- on accède enfin à la représentation de l'objet par un processus d'appariement utilisant les Géons et leurs relations spatio-temporelles.

C'est donc un modèle de type bottom-up dans lequel les relations qu'entretiennent les Géons entre eux sont des éléments cruciaux de l'identification de l'objet.

Une autre organisation de la perception des formes par le système visuel est proposée par Bonnet (Bonnet, 1995). Le traitement de l'information se divise en deux grandes étapes :

- on retrouve un traitement initial ascendant au cours duquel s'effectue l'extraction en parallèle des primitives visuelles. Il est suivi d'une structuration de ces données selon des configurations permettant une ségrégation de la figure et du fond. Ce traitement est automatique donc non influençable par les informations d'ordre supérieur (top-down). À la fin de cette étape, une représentation perceptive transitoire est élaborée.

- c'est au niveau de cette représentation transitoire que sont combinées les informations sensorielles et les connaissances préalables issues de la mémoire à long terme pour élaborer la représentation finale univoque de l'objet.

Toutes les expériences supportant ces théories et modèles ont été effectuées sur des images pictographiques simples (objets ou formes), en noir et blanc ou en couleur, sur un fond uniforme. À ce titre, elles peuvent difficilement rendre compte des processus perceptifs impliqués dans la vision de scènes naturelles complexes et variées. Certes, les phénomènes perceptifs sont tellement complexes qu'il est souvent nécessaire de simplifier au maximum la stimulation utilisée pour mieux les cerner. Il ne faut pas pour autant négliger les informations contextuelles apportées par l'environnement des images naturelles.

Informations contextuelles

Le contexte peut se définir comme l'ensemble des aspects d'une situation qui ne sont pas directement sollicités dans la réponse mais qui l'influencent. Dans la présente étude, la notion de contexte fait référence aux relations spatiales qui existent entre les différents éléments de la scène naturelle. En effet, les diverses entités qui composent une scène ne forment pas une collection inorganisée mais sont reliées entre elles par des relations spatiales déterminées par leur nature. Le poids informatif du contexte d'une image sera d'autant plus affecté que les relations, dites "logiques", entre les objets seront perturbées. Ces relations logiques sont issues de l'expérience perceptive et donc dépendent à la fois des connaissances du monde et de certaines règles de bonne organisation spatiale.

Les informations contextuelles relèvent donc directement de la voie top-down.

Études portant sur l'identification de scènes naturelles

Sur des images naturelles, Biederman montre que nos connaissances du monde et nos attentes accélèrent et rendent plus efficace notre perception (Biederman, 1972 ; Biederman, Glass & Stacy, 1973). Cette affirmation est basée sur la comparaison des performances obtenues par des sujets devant identifier une scène selon que l'image est présentée normalement ou avec un contexte mélangé. La modification du poids informatif du contexte de l'image se répercute sur la qualité et la rapidité d'analyse de la scène, et cet effet est obtenu pour des temps de présentation très courts (50 à 100ms). Ces conclusions se vérifient à la fois dans une tâche d'identification et dans une tâche de recherche d'un objet-cible dans une scène naturelle. De même, la violation des règles qui

gèrent la cohérence des relations spatiales entre les objets de la scène affecte la perception des sujets (Biederman, Mezzanotte & Rabinowitz, 1982).

Ces résultats sont à rapprocher de ceux obtenus par Miller, Heise & Lichten (1951) sur l'intelligibilité des mots dans le bruit : un mot présenté dans une phrase bien formée, dont le contexte est cohérent, est plus intelligible que s'il est présenté parmi des mots aléatoirement choisis, c'est à dire privé de tout contexte.

Tout comme la perception d'une phrase est déterminée par les relations sémantiques et syntaxiques qui existent entre les mots, il est nécessaire de spécifier les relations entre les différents éléments d'une scène visuelle complexe (Biederman, 1981).

Il reste difficile de définir le rôle que joue le contexte dans la perception de scènes naturelles. En se basant sur des critères temporels, Thorpe souligne que même si le contexte peut faciliter les processus d'identification, comme le montrent les expériences de Biederman (1972), l'identification d'images complexes doit être possible avec des systèmes dans lesquels le contexte ne joue pas un rôle important (Thorpe, 1988). L'étude de Biederman sur la répercussion de violations des règles d'organisation spatiale d'une scène naturelle aboutit elle-même à une conclusion comparable spécifiant qu'en un seul coup d'œil on a accès aux relations sémantiques, physiques et à l'identification des objets (Biederman & col., 1982).

Les études supportant les modèles théoriques reposent pour la plupart sur une tâche de recherche visuelle ou d'identification. Les rôles respectifs des différentes voies de traitement restent mal définis dans le cadre d'une tâche de reconnaissance.

Modèle d'organisation du système visuel dans une tâche de recherche visuelle ou d'identification

L'étude proposée vise à évaluer dans quelle limite le sujet peut se passer de ses connaissances, à quel moment il utilise les informations contextuelles dans une tâche de reconnaissance et de localisation d'un objet dans une scène visuelle naturelle. Pour cela, on suppose valide pour la tâche de reconnaissance d'une forme un modèle d'organisation du système visuel décrit par Bonnet pour l'identification (Bonnet, 1995). Ce schéma général décrit un système de traitement de l'information visuelle où la reconnaissance d'un objet dépend de l'interaction entre un flux ascendant véhiculant les informations sensorielles et un flux descendant provenant de la Mémoire à Long Terme et apportant les informations de connaissance sur l'organisation spatiale du monde, issue de l'expérience.

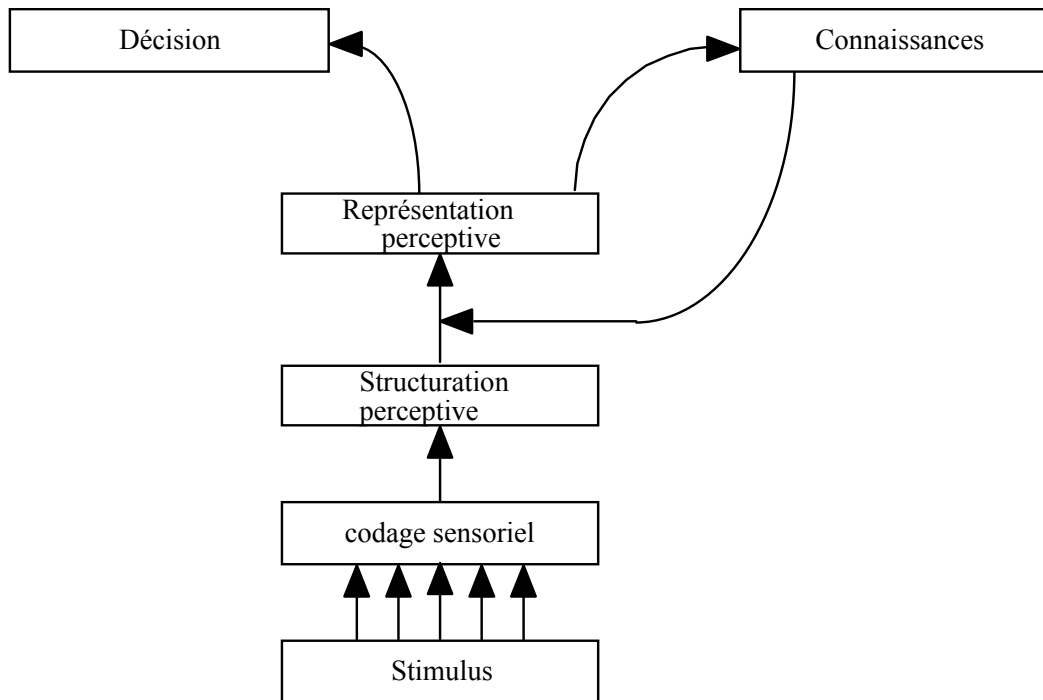


Figure 1 : schéma général d'organisation du système visuel

Si l'on considère la voie de traitement ascendante, il semble que l'on puisse se passer de l'aide apportée par les informations de contexte dans le cadre d'une tâche de reconnaissance ou d'identification d'un objet dans une scène naturelle et dans des conditions de perception dites "normales" ou "habituelles" (Thorpe, 1988). La perception de l'espace s'effectue habituellement à partir du pouvoir d'analyse et de résolution spatio-temporelle des capteurs sensoriels. Cependant, ces conditions ne sont pas immuables et peuvent être dégradées de façon naturelle (brouillard, éblouissement) ou artificielle (lunettes teintées, dispositifs optiques variés, amétropie mal compensée). En se replaçant dans le cadre du schéma général vu précédemment, il s'avère nécessaire d'évaluer les capacités du système visuel vis-à-vis de ces conditions de signaux sensoriels dégradés. En considérant le niveau des informations afférentes, les facteurs de variation sont les suivants : les fréquences spatiales, le contenu chromatique, le mouvement, le contraste, l'orientation, la disparité binoculaire, etc. La présente expérimentation s'adresse plus particulièrement au contenu fréquentiel et chromatique de la stimulation et indirectement à l'influence du contraste de luminance.

Informations sensorielles

Certains auteurs ont tenté de déterminer l'importance relative des différentes fréquences spatiales et des informations de couleur dans la reconnaissance et l'identification d'objets. Sergent (1986), dans une tâche d'identification des visages, suggère que les basses fréquences sont destinées à la catégorisation visage/non visage, les moyennes fréquences permettent la distinction homme/femme et les hautes fréquences sont réservées à l'identification proprement dite. Pour Ginsburg (1978), Harmon (1973) et Sergent (1986), le rôle des informations grossières, véhiculées par les basses fréquences spatiales, est crucial puisqu'elles permettent la classification initiale

de la catégorie dans laquelle se situe l'objet à identifier. Les informations de détails, contenues dans les hautes fréquences spatiales, ne participent alors qu'à la phase terminale du processus d'identification (Parker, Lishman & Hughes, 1996).

Dans la littérature, la plupart des études évaluant l'influence du contenu fréquentiel spatial sur la perception des formes, portent sur les stades précoces de traitement de l'information visuelle, comme la Fonction de Sensibilité aux Contrastes, ou étudient des fonctions de haut niveau mais avec des stimulus simples ou isolés. Rares sont les études cumulant des images de scènes naturelles, une tâche complexe de reconnaissance ou d'identification et des dégradations fréquentielles.

Dans ce domaine, il faut souligner les travaux d'Oliva & Schyns (1994) combinant dans la stimulation visuelle des images naturelles filtrées sur des gammes disjointes de fréquences spatiales (images hybrides). Les sujets identifient préférentiellement la scène dont le contenu en fréquences spatiales est centré sur des fréquences plus basses. Oliva et Schyns suggèrent que l'analyse d'une scène naturelle commence par une activation rapide des représentations de scènes compatibles avec la structure des basses fréquences de la stimulation. Le produit de cette activation précoce agit ensuite comme contrainte sur les informations plus tardives de détails véhiculées par les hautes fréquences spatiales. Selon ces auteurs, les basses fréquences sont cruciales puisqu'elles vont véhiculer directement l'information concernant les relations spatiales entre les différents éléments de la scène et cette information sera la base de toute l'analyse de l'image.

Ces résultats vont dans le sens communément admis d'un traitement plus rapide des informations contenues dans les basses fréquences spatiales. Contrairement à Ginsburg (1978), Harmon (1973), Sergent (1986) ou Oliva et Schyns (1996), certains auteurs dont Marr (1982), Biederman (1987), Lowe (1987) estiment que cette prévalence n'a aucun rôle spécifique dans des opérations de haut niveau comme l'identification. Biederman & Ju (1988) ont montré que l'on nomme aussi rapidement une image grossière dessinée avec des traits noirs et blancs qu'une photographie détaillée et en couleur.

Comme le précisent Oliva & Schyns, il faut relativiser toutes ces théories. Il paraît difficilement concevable de vouloir attribuer un rôle particulier et définitif à chaque type d'information. L'importance relative des informations doit nécessairement varier en fonction de la tâche demandée. Cet aspect est illustré dans une étude d'Uttal, Baruch & Allen (1995) comparant la tâche de discrimination à celle de reconnaissance par "le phénomène Harmon & Julesz". Harmon & Julesz (1973) ont montré que des images décomposées en blocs sont plus facilement reconnaissables quand elles sont filtrées de leurs hautes fréquences spatiales (par des filtres passe-bas). Cette constatation a conduit Harmon & Julesz à proposer que les bandes de hautes fréquences apportées par la décomposition en blocs de l'image masquent les basses fréquences spatiales porteuses de l'information nécessaire pour la reconnaissance. Le filtrage des hautes fréquences spatiales perturbatrices améliore donc la reconnaissance. D'après les résultats obtenus par Uttal & col., il semble que la tâche soit aussi importante que les caractéristiques spectrales du stimulus. Le phénomène Harmon et Julesz est confirmé dans une tâche de reconnaissance mais ne l'est pas dans une tâche de discrimination. Uttal & col. en concluent qu'il existe une différence qualitative plutôt que quantitative entre ces deux tâches, à savoir qu'elles n'utilisent pas le même type d'information. Alors que la reconnaissance est fondée sur les informations globales (basses fréquences spatiales) de l'image, la discrimination s'intéresse plutôt aux informations locales

(hautes fréquences spatiales). Selon ces auteurs, le spectre fréquentiel est seulement l'un des facteurs saillants affectant la perception, les autres facteurs étant : la tâche, le stimulus, l'ordre, l'organisation. Donc, pour Uttal & col., il est difficile d'envisager d'une part un modèle universel de la perception et d'autre part un modèle strictement bottom-up et automatique.

Informations colorées

La plupart des études portant sur l'effet de la couleur dans la perception de scènes naturelles s'accordent d'une part sur le fait que les informations chromatiques servent essentiellement à la segmentation de la scène, c'est à dire à distinguer l'objet de son fond, et d'autre part sur le fait que ce sont les basses fréquences dans les canaux chromatiques qui sont les plus utiles pour la reconnaissance (Gegenfurtner, 1996 ; Wurm, Legge, Isenberg & Luebker, 1993 ; Biederman & Ju, 1988). Les différences entre les auteurs apparaissent à propos de l'effet facilitateur ou non de la couleur sur la reconnaissance ou l'identification de scènes naturelles. Gegenfurtner (1996) et Wurm & col. (1993) montrent respectivement pour une tâche de reconnaissance et d'identification que la couleur améliore les performances des sujets. Au contraire Biederman & Ju (1988) constatent qu'il n'y a pas de différence de temps de réponse entre des photographies en couleur et des dessins en noir et blanc. Il faut néanmoins pondérer ces résultats par le fait qu'aucun contrôle de l'isoluminance entre les présentations avec et sans couleur n'est effectué dans l'étude de Biederman & Ju. Les résultats montrant une amélioration de la tâche due à la couleur sont pour leur part étroitement fonction des objets utilisés.

De plus, les comparaisons sont faites entre des images colorées et des images en noir et blanc, ce qui ne correspond en rien, pour ces dernières, à une quelconque condition de vision naturelle. En effet, les cas d'achromatisme étant excessivement rares, les seules conditions de détérioration du sens coloré porteront soit sur des cas pathologiques de dyschromatopsie ou de rétinopathie, soit sur des cas occasionnés par le port de filtres colorés, soit enfin sur des cas de fort éblouissement. Quelles que soient les conditions, des perturbations peuvent intervenir mais seulement dans des tâches très particulières mettant en jeu des comparaisons ou des recherches spécifiques de couleurs.

Hypothèses de l'étude

L'étude présentée dans ce document se situe à la jonction de flux d'informations ascendants et descendants. Elle combine des facteurs pouvant faire varier le contenu informationnel du contexte et le contenu physique de l'information sensorielle, affectant donc séparément l'une ou l'autre des deux voies de traitement.

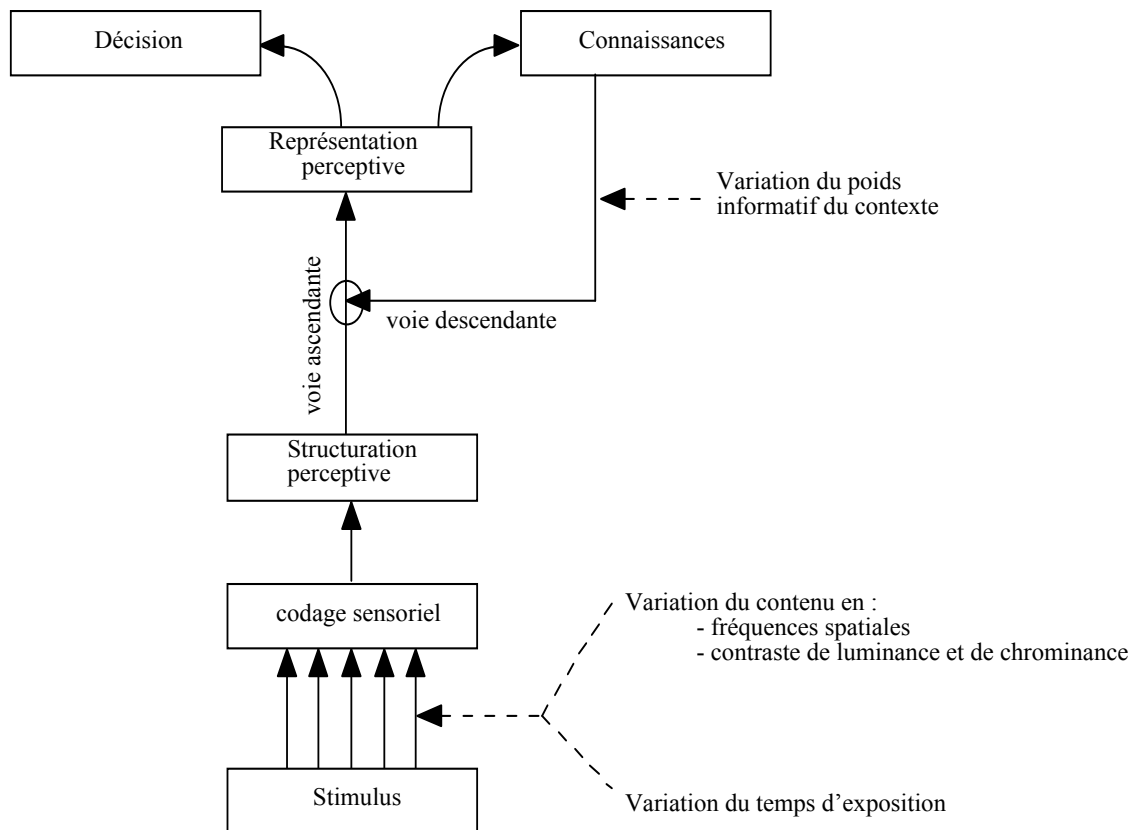


Figure 2 : points d'impact des facteurs de variation

L'étude se propose d'évaluer si le même schéma d'organisation générale, mis en évidence au niveau de la tâche d'identification, peut être reconduit pour la tâche de reconnaissance analysée dans cette expérience. Dans le cas où le modèle est validé, il s'agit d'évaluer le poids respectif des informations sensorielles ascendantes et des informations contextuelles descendantes dans la performance générale des sujets ainsi que l'évolution de ces poids respectifs au fur et à mesure de la répétition des différentes conditions de stimulation.

3- PROTOCOLE EXPÉRIMENTAL

3-1- Matériel

Les images tests sont présentées aux sujets sur un moniteur couleur Hitachi de diagonale 21 pouces. Le séquençage des présentations et le recueil des données sont gérés par une station de travail "Silicon Graphics".

Le moniteur est muni d'un écran tactile permettant de recueillir les réponses des sujets. Les sujets sont installés à 50 cm de l'écran, la tête appuyée sur une mentonnière et la main posée à 40 cm de l'écran sur la barre d'espacement du clavier. L'expérimentation est menée dans une salle sans autre source lumineuse que le moniteur pour éviter tout reflet sur l'écran du moniteur perturbant la tâche du sujet.

3-2- Stimulations

Les stimulations visuelles utilisées sont des images naturelles, issues de la base de données COREL, et représentent des scènes de la vie quotidienne : champêtres, citadines ou d'intérieur (voir Annexe).

Huit images ont été sélectionnées selon des critères de taille, de localisation, de couleur et d'environnement des objets. L'objet cible est situé dans la case centrale d'une partition équilibrée de l'image formée de 5 cases en largeur sur 3 cases en hauteur (matrice virtuelle de la figure 3). D'un point de vue chromatique, l'objet cible et son fond sont ou non dans la gamme chromatique la plus absorbée par le filtrage chromatique retenu (c'est-à-dire la bande de longueur d'onde centrée sur le vert 525 nm : voir la courbe d'absorption en annexe) ; il existe 2 images pour chaque combinaison chromatique "objet x fond". Enfin, les images sont équilibrées : des objets, ou éléments sémantiques, sont répartis sur toute la surface des images.

Chaque image est définie par 1705 x 1152 pixels. Sur la matrice virtuelle à 15 cases, un masque de sélection correspondant à 6 cases (3 x 2) permet d'extraire 6 sous-images de taille 1024 x 768 pixels qui forment la base d'images initiales affichées à l'écran (figure 3 et voir exemple en fin d'annexe). Pour chacune de ces sous-images, l'objet cible a une position différente sur l'écran. Compte tenu de la présence régulière d'éléments dans les images de départ, on considère que l'information sémantique apportée par l'environnement ne varie pas d'une sous-image à l'autre.

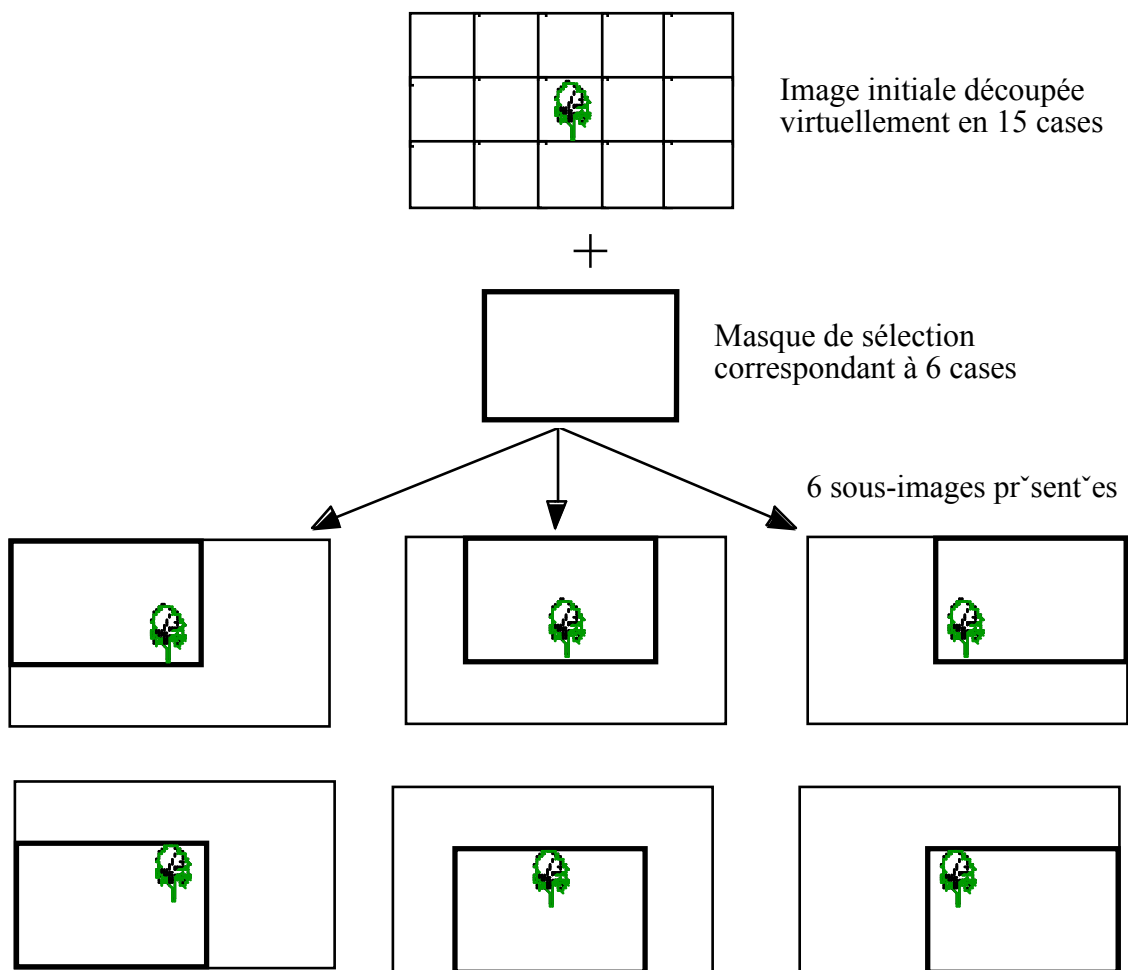


Figure 3 : Méthode de sélection des images présentées

Dans les conditions d'observation, l'écran du moniteur sous-tend un angle de 39 degrés en horizontal par 31,5 degrés en vertical. L'image est affichée en position centrale et sous-tend un angle de 31 x 23,7 degrés. Le reste de l'écran et la totalité de l'écran en l'absence d'image sont formés par une plage uniforme, de coordonnées colorimétriques : $x=0,276$ et $y=0,311$ (correspondant à une répartition uniforme de la luminance sur les trois types de luminophores) et de luminance moyenne 35 cd.m^{-2} . Cette luminance correspond à la luminance moyenne sur l'ensemble des images. La luminance moyenne des différentes images initiales varie de $21,6 \text{ cd.m}^{-2}$ à $40,3 \text{ cd.m}^{-2}$. Les stimulations sont donc de type photopique. Les coordonnées colorimétriques ne peuvent pas être exhaustivement décrites en raison de la large gamme de couleurs distribuée sur l'ensemble des images et exploitant l'étendue des possibilités de génération du moniteur.

3-3- Variables expérimentales

Quatre facteurs sont contrôlés dans cette expérimentation pour étudier leur retentissement sur la reconnaissance d'objets cibles. Ce sont :

- le contenu fréquentiel de la stimulation ;
- le contenu chromatique de la stimulation ;
- le contenu contextuel de la stimulation ;
- le temps de présentation de la stimulation.

3-3-1- Variation du contenu fréquentiel

À partir de l'image initiale, trois niveaux de filtrage en fréquences spatiales ont été retenus. Les filtrages sont obtenus par décomposition de l'image initiale à l'aide de l'analyse en ondelettes (Mallat, 1989) avec un profil de base des fonctions analysantes en différence de deux gaussiennes. Cette méthode d'analyse d'images a l'avantage d'extraire à chaque échelle un contenu fréquentiel spatial de l'image comparable à la sélectivité en fréquences spatiales du système visuel (Plantier, Menu, 1993). Les images expérimentales résultent ensuite d'une reconstruction partielle fonction des niveaux de filtrage désirés. Le premier niveau de filtrage élimine les fréquences spatiales supérieures à 6 cycles/degré (cpd), le second niveau de filtrage les fréquences supérieures à 3 cycles/degré, et le troisième niveau de filtrage les fréquences supérieures à 1,5 cycles/degré (figure 4). Les différents niveaux de filtrage correspondent à une amputation de plus en plus étendue du spectre fréquentiel spatial des images initiales dans le domaine des hautes fréquences spatiales. Ces dégradations du contenu fréquentiel de l'image suppriment de l'information concernant essentiellement les détails fins et rendent les images de plus en plus floues au fur et à mesure que s'élève le niveau de filtrage.

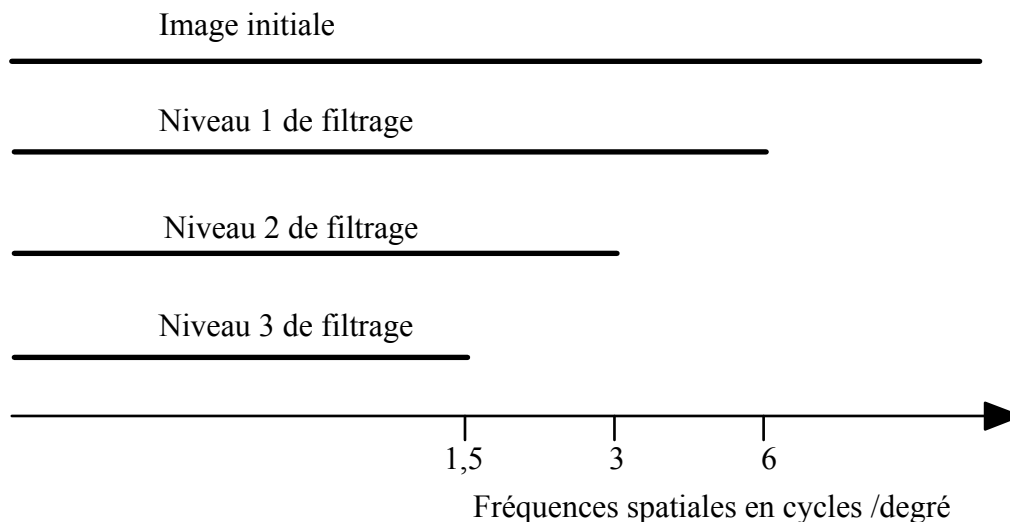


Figure 4 : Contenu fréquentiel résultant des filtrages en fréquences spatiales

3-3-2- Variation du contenu chromatique

La dégradation chromatique est effectuée en attribuant des coefficients d'atténuation différents aux 3 composantes colorées (Rouge, Vert, Bleu) des images. Ces coefficients sont définis à partir des paramètres spectraux d'un système optique filtrant destiné à la protection oculaire. Les coefficients d'atténuation appliqués sont :

- 0,1 pour le Rouge,
- 0,07 pour le Vert,
- 0,06 pour le Bleu.

Le filtrage chromatique s'applique sur les trois composantes et entraîne une baisse d'environ 90% des niveaux de luminance (tableau 1).

Tableau 1 : Variation de la luminance induite par le filtrage chromatique

	Luminance moyenne de l'image sans filtrage en fréquences spatiales (cd.m ⁻²)	
	Sans filtrage chromatique	Avec filtrage chromatique
Feu	26,7	2,7
Poubelle	29,4	2,28
Bonnet	33,1	2,6
Chalet	30,9	2,4
Grenouille	21,6	1,7
Chapeau	22,1	1,8
Bouteille	26	2,1
Jeep	40,3	3,2

3-3-3- Variation du contenu contextuel

Le principe de modification du contenu contextuel d'une image est issu d'une étude de Biederman (1972). Il repose sur un découpage de l'image en 6 cases (3 x 2) et

sur un mélange aléatoire de ces 6 cases. Cette transformation de l'image rompt les relations spatiales cohérentes entre les différents éléments de la scène (figure 5). Le poids informatif du contexte varie entre une image structurée et son image complémentaire déstructurée. Afin de pouvoir comparer les performances obtenues pour ces 2 types d'images, l'objet cible est systématiquement situé au même endroit dans l'image structurée et déstructurée. De plus, les cases ayant un côté commun avec celle de l'objet cible sont automatiquement déplacées dans l'image déstructurée.

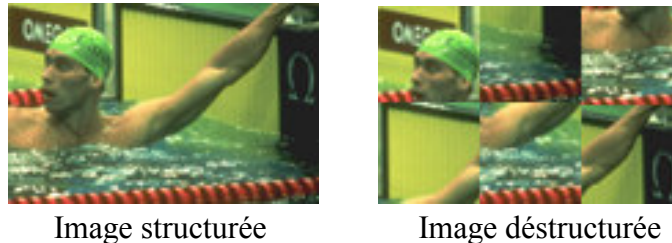


Figure 5 : Modification du contenu contextuel

3-3-4- Contrôle du temps de présentation

Quatre temps de présentation des images ont été choisis : 100, 200, 400 et 800ms. Cette contrainte sur le temps de présentation est à la fois un moyen de mesure du temps nécessaire pour que la tâche soit accessible et également une contrainte sur la prise d'information visuelle.

3-3-5- Synthèse et notation des différentes variables

Pour chaque image initiale, il est obtenu :

- 4 niveaux de filtrage en fréquences spatiales :
 - image initiale, notée fs0,
 - 1er niveau sans les fréquences supérieures à 6 cpd, noté fs1,
 - 2ème niveau sans les fréquences supérieures à 3 cpd, noté fs2,
 - 3ème niveau sans les fréquences supérieures à 1,5 cpd, noté fs3 ;
- 2 niveaux de filtrage chromatique :
 - image sans filtrage chromatique,
 - image avec filtrage chromatique ;
- 2 niveaux de contexte :
 - image structurée avec des relations cohérentes entre les différents éléments de la scène,
 - image déstructurée avec une perturbation des relations cohérentes entre les éléments de la scène ;
- 4 Temps de présentation : 100, 200, 400 et 800 ms.

Enfin, la position de l'objet sur l'écran est également contrôlée de façon que l'objet apparaisse une fois à gauche, une fois au centre et une fois à droite de l'écran pour chaque condition expérimentale.

Pour chaque image, toutes ces conditions sont croisées et permettent de générer 192 configurations (4 niveaux de filtrage en fréquence spatiales x 2 niveaux de filtrage chromatique x 2 niveaux de contexte x 4 temps de présentation x 3 positions). Les 8 images initiales donnent donc 1536 présentations différentes (8 x 192 configurations par image).

3-4- Déroulement temporel

3-4-1- Organisation générale de l'expérimentation

L'expérience se découpe en plusieurs séances. Les 1536 présentations sont affichées en ordre aléatoire et différent selon les sujets. Elles sont réparties en 4 blocs équilibrés de présentations. Un bloc fait l'objet d'une séance expérimentale au cours de laquelle 384 images sont présentées.

Chaque sujet effectue 3 répétitions des 1536 présentations dans le même ordre (i.e. 12 séances par sujet).

L'enregistrement des données est précédé d'une phase de familiarisation avec le dispositif expérimental et avec la tâche. Cette familiarisation comporte :

- 1- une présentation générale de l'expérimentation au sujet ;
- 2- une explication orale du déroulement temporel d'une présentation et de la tâche à effectuer, avec l'aide d'un schéma synoptique ;
- 3- une présentation visuelle des 8 objets-cibles, isolés de leur environnement ;
- 4- une séance d'entraînement composée d'au moins 50 présentations dans des conditions expérimentales variées ;

3-4-2- Déroulement d'une présentation

- Évocation sémantique : le nom de l'objet-cible est affiché pendant une seconde.
- Le nom est remplacé par un carré de fixation (de 0,6 deg.), au centre de l'écran. Il permet au sujet de centrer son regard avant chaque présentation d'image. Le carré de fixation reste affiché jusqu'à ce que le sujet déclenche la poursuite de la séquence en appuyant sur la barre d'espace du clavier de l'ordinateur.
- L'image test est affichée pendant une durée correspondant à la condition expérimentale en cours (100, 200, 400 ou 800ms).
- À l'issue de la présentation de l'image test, une grille est affichée à l'écran qui matérialise 6 cases (3x2) théoriques de l'image et parmi lesquelles le sujet doit désigner manuellement celle où il a perçu l'objet cible. La consigne est de pointer le plus rapidement possible l'emplacement de l'objet cible, tout en privilégiant la qualité de la réponse.

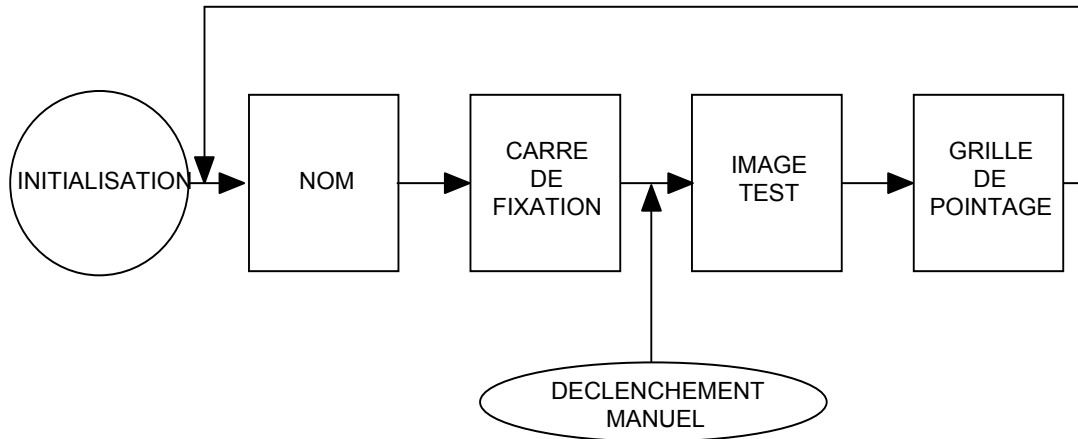


Figure 6 : Déroulement d'une séquence

À chaque essai sont enregistrés la case de pointage et le temps de réponse défini par l'intervalle séparant l'appui sur la barre d'espacement et le contact sur l'écran tactile.

3-5- Sujets

L'expérience a été menée sur 5 sujets, 2 femmes et 3 hommes, âgés de 23 à 29 ans. Chaque sujet est emmétrope ou rendu tel grâce à une compensation optique (lunettes ou lentilles de contact). Dans les conditions de l'expérimentation, tout sujet a une acuité visuelle minimale de 10/10 sur une échelle d'acuité de type Monoyer. La sensibilité des sujets aux différentes fréquences spatiales est évaluée par une mesure de la fonction de sensibilité aux contrastes qui doit être normale sur toute l'étendue du spectre. Enfin, les sujets ne doivent commettre aucune erreur lors de l'évaluation de leur sensibilité chromatique par le test d'Ishihara.

3-6- Paramètres recueillis

Les paramètres recueillis et analysés sont :

- le taux d'erreurs exprimé par le pourcentage des présentations où le sujet désigne un objet différent de l'objet recherché par rapport à l'ensemble des présentations effectuées dans les mêmes conditions expérimentales ;
- le temps de réponse dans le cas où la réponse est correcte et désigne l'objet à reconnaître dans l'image.

3-7- Analyse des résultats

L'analyse conjointe du taux d'erreurs et du temps de réponse lors de réponses avec désignation effective de la cible permet une approche globale de la performance des sujets. Le taux d'erreurs concerne les réponses incorrectes, soit un échec dans l'exécution de la tâche. Le temps de réponse des réponses correctes rend compte de la difficulté dans l'exécution réussie de la tâche.

Les données sont analysées avec une analyse de variance (ANOVA) qui permet d'évaluer les paramètres recueillis ou données quantitatives (temps de réponse et taux

d'erreurs) en fonction des données qualitatives citées précédemment (contenu fréquentiel, contenu chromatique, contenu contextuel et temps d'exposition).

4- RÉSULTATS

4-1- Taux d'erreurs

Au cours des 23040 présentations d'images, 1568 erreurs de pointages de la cible ont été effectuées (soit 6,8% des présentations). L'analyse des données porte sur la proportion d'erreurs de pointage obtenues dans les différentes conditions expérimentales par chacun des sujets. Pour chacune des 512 conditions de présentation d'images (8 images x 4 filtrages en fréquences spatiales x 2 filtrages chromatiques x 2 niveaux de contexte x 4 temps de présentation) on évalue un taux d'erreurs correspondant au pourcentage de désignation incorrecte de la cible au cours des 3 répétitions des 3 positions de cibles envisagées. On dispose donc de 2560 taux d'erreurs pour les 5 sujets de l'expérimentation (512 x 5). Les résultats sont exprimés en moyenne de taux d'erreurs et son erreur standard.

4-1-1- Taux d'erreurs sur les images d'origine

Si l'on considère les images n'ayant subi aucun traitement (ni filtrage en fréquences spatiales, ni filtrage chromatique, ni désorganisation de l'image), seulement 5 erreurs ont été commises sur les 640 présentations, soit un taux d'erreurs moyen de 0,78%. Les différentes images sont comparables entre elles dans ces conditions de base, et l'on peut considérer que la tâche de reconnaissance-localisation d'objet cible est accessible aux observateurs avec ces images pour des temps de présentation supérieurs ou égaux à 100ms.

4-1-2- Variabilité liée aux sujets et aux images

Le taux d'erreurs moyen varie de $5,95 \pm 0,63$ à $8,12 \pm 0,77\%$ entre les sujets (figure 7) mais ces différences ne sont pas significatives au vu de la dispersion des données (effet Sujet : $F^4_{2555}=1,75$; $p=0,14$). Les résultats des 5 sujets seront considérés globalement.

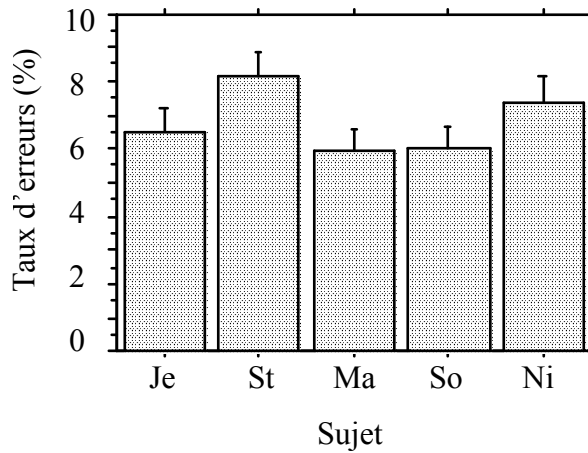


Figure 7 : taux d'erreurs moyen des sujets

En revanche, le taux d'erreurs dépend de l'image présentée. Sa valeur moyenne varie de $0,48 \pm 0,16\%$ pour la reconnaissance du chapeau à $16,11 \pm 1,27\%$ pour celle de la jeep (figure 8). L'ensemble des images se répartit en 3 groupes : celui des images où les objets sont faciles à reconnaître (bonnet, chapeau, bouteille), celui des images où l'objet est difficile à reconnaître (poubelle, jeep), et un groupe intermédiaire (feu, chalet, grenouille). La différence des taux d'erreurs entre les images est significative (effet Image : $F^7_{2552}=55,1$; $p < 0,0001$).

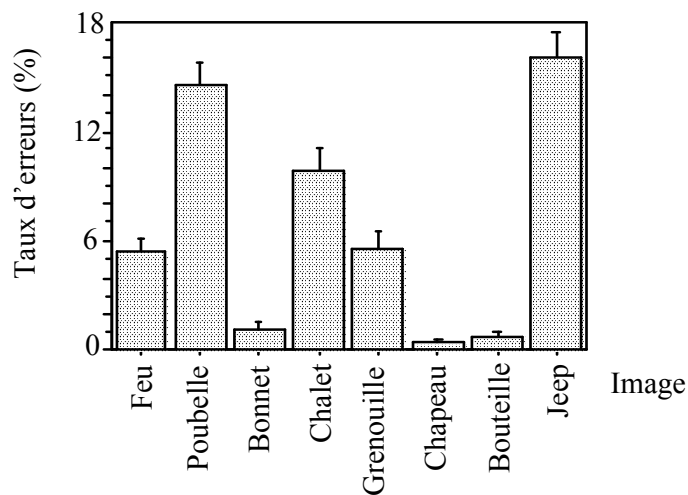


Figure 8 : Taux d'erreurs moyen sur les différentes images.

L'échantillon d'images retenu pour l'expérimentation sur des critères photolorimétriques et d'organisation uniformément distribuée des objets de la scène dans l'image, s'avère représentatif des images naturelles où les objets sont plus ou moins faciles à reconnaître. D'un point de vue méthodologique, cette variabilité conduit à envisager l'influence de la particularité des diverses images sur les résultats obtenus.

4-1-3- Facteurs de variation sur l'ensemble des données

4-1-3-1- Effet de la dégradation de l'image

4-1-3-1-1- Dégradation du contenu en fréquences spatiales

Le taux d'erreurs augmente progressivement au fur et à mesure que l'on ampute le spectre en fréquences spatiales des images (figure 9). Le taux d'erreurs passe ainsi de $1,77 \pm 0,23\%$ dans le cas des images non filtrées, à $16,23 \pm 0,95\%$ dans le cas des images ayant subi le filtrage maximum en fréquences spatiales. L'effet du filtrage est surtout manifeste lorsque l'on enlève les fréquences spatiales supérieures à 1,5cpd (fs3), puisque le taux d'erreurs est alors quasiment triplé. L'influence du filtrage sur le taux d'erreurs est statistiquement significative (effet Fréquence spatiales : $F^3_{2048}=422,2$; $p < 0,0001$).

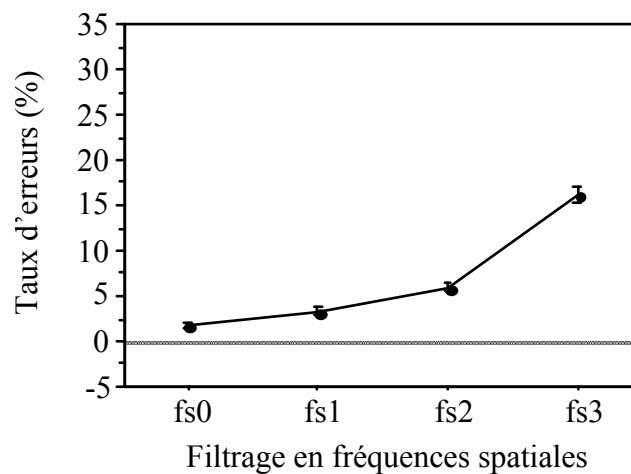


Figure 9 : Effet du filtrage en fréquences spatiales.

Si toutes les images présentent une augmentation systématique du taux d'erreurs avec l'application des filtres successifs, la performance de reconnaissance-localisation n'est pas atteinte dans les mêmes proportions (figure 10). Les images faciles d'interprétation résistent bien à l'effet du filtrage. Certaines images par contre, comme le feu ou la jeep, ont une augmentation drastique de leur taux d'erreurs lors de l'application du troisième filtre. Ce niveau de filtrage supprime une information permettant la discrimination de l'objet recherché par rapport aux autres éléments de la scène. Enfin, l'image contenant la poubelle engendre une altération régulière de la performance sur chacun des niveaux de filtrage.

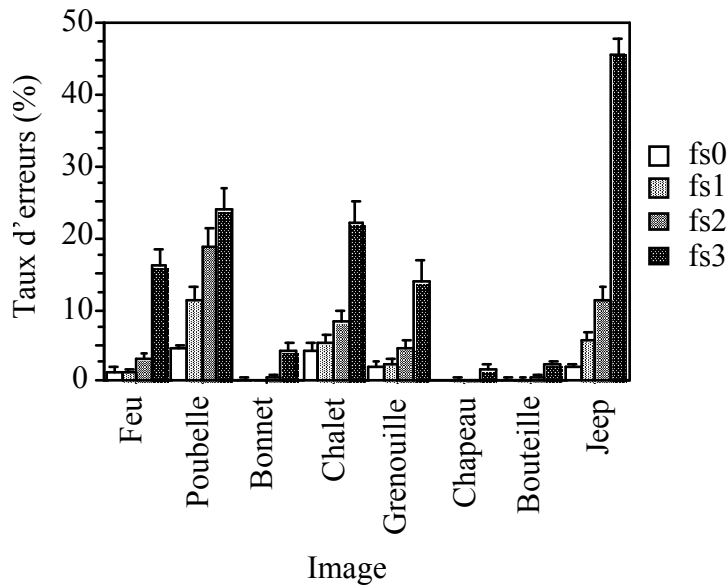


Figure 10 : Effet du filtrage fréquentiel sur le taux d'erreurs moyen obtenu avec les différentes images.

4-1-3-1-2- Dégradation du contenu chromatique

Le taux d'erreurs dépend également de la disponibilité des informations colorées : il augmente d'un facteur 4 lorsque les images présentées subissent un filtrage centré sur les grandes longueurs d'ondes (figure 11), passant de $2,79 \pm 0,26\%$ en l'absence de filtre chromatique à $10,77 \pm 0,56\%$ en présence du filtre. Cette altération de la performance est d'autant plus marquée que le contenu en fréquences spatiales des images est appauvri. L'analyse statistique des données confirme un effet Filtre chromatique ($F^1_{2048}=626,9$; $p < 0,0001$) et une interaction Fréquences spatiales x Filtre chromatique ($F^3_{2048}=90,2$; $p < 0,0001$).

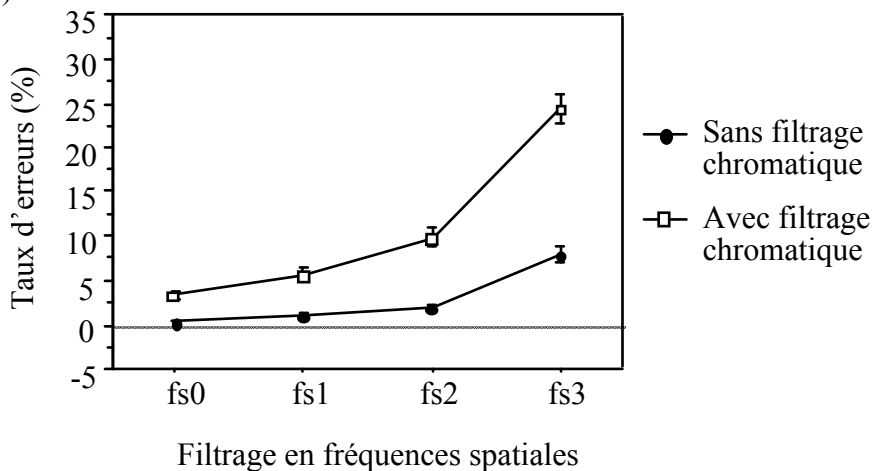


Figure 11 : Interdépendance des filtres fréquentiels et chromatique sur le taux d'erreurs moyen.

Certaines images subissent plus largement l'effet du filtrage chromatique (figure 12). C'est principalement le cas des images où l'application du filtrage supprime un contraste de luminance entre un objet et son fond (le chalet sombre sur fond de prairie verte) ou diminue notablement le niveau lumineux des images (la grenouille verte sur fond vert). Dans certains cas, par contre, la variation du taux d'erreurs est moins nette (images contenant la poubelle ou la jeep) en raison de l'influence majeure des autres

facteurs de variation sur ces images, ayant un taux d'erreurs élevé dans la condition non filtrée chromatiquement.

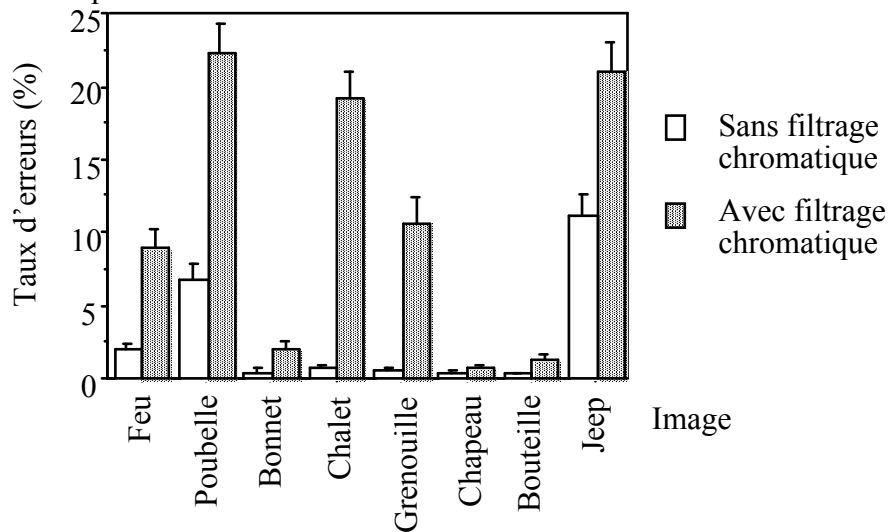


Figure 12 : Effet du filtrage chromatique sur le taux d'erreurs pour les différentes images.

4-1-3-2- Effet du temps de présentation

Toutes conditions confondues, le taux d'erreurs moyen augmente au fur et à mesure que l'on réduit le temps de présentation des images, avec un facteur 5 entre les durées d'affichage extrêmes. Il n'est que de $2,47 \pm 0,38\%$ lorsque les images sont affichées pendant 800ms, mais atteint $14,10 \pm 0,90\%$ lorsque l'affichage est réduit à 100ms. Cet effet est d'autant plus marqué que les images présentées ont leur contenu physique appauvri en termes de fréquences spatiales et/ou de contenu chromatique (figures 13 et 14). L'analyse statistique des données confirme un effet Temps de présentation ($F^3_{2048}=258,8$; $p<0,0001$) et des interactions Temps de présentation x Fréquences spatiales ($F^3_{2048}=27,3$; $p<0,0001$) Temps de présentation x Filtre chromatique ($F^3_{2048}=130,2$; $p<0,0001$).

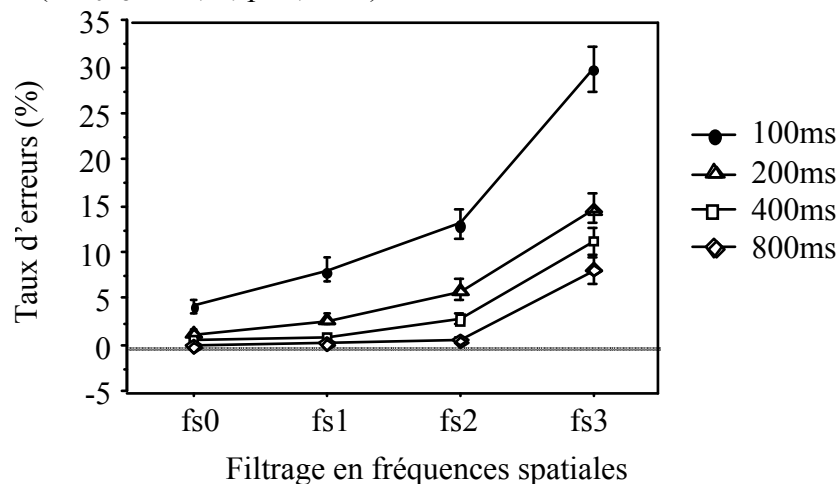


Figure 13 : Interdépendance du filtrage fréquentiel et de la contrainte temporelle sur le taux d'erreurs moyen.

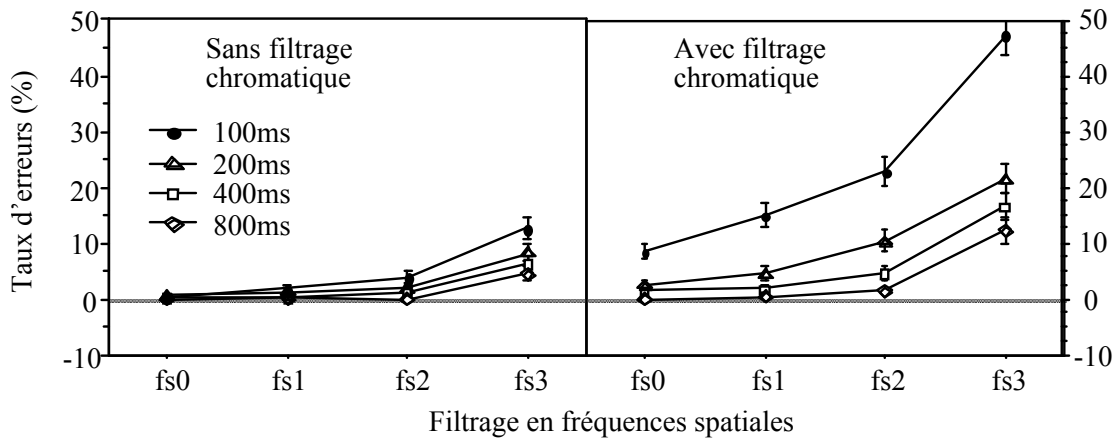


Figure 14 : Effet conjoint des trois restrictions dans la prise d'information

Pour toutes les images, le taux d'erreurs diminue lorsqu'on allonge le temps de présentation (figure 15). Le taux d'erreurs devient négligeable lorsque les images sont affichées pendant 800ms, de l'ordre de $1,17 \pm 0,21\%$, sauf dans le cas de l'image contenant la jeep où le taux d'erreurs est encore de $10,52 \pm 2,42\%$ dans ces conditions là.

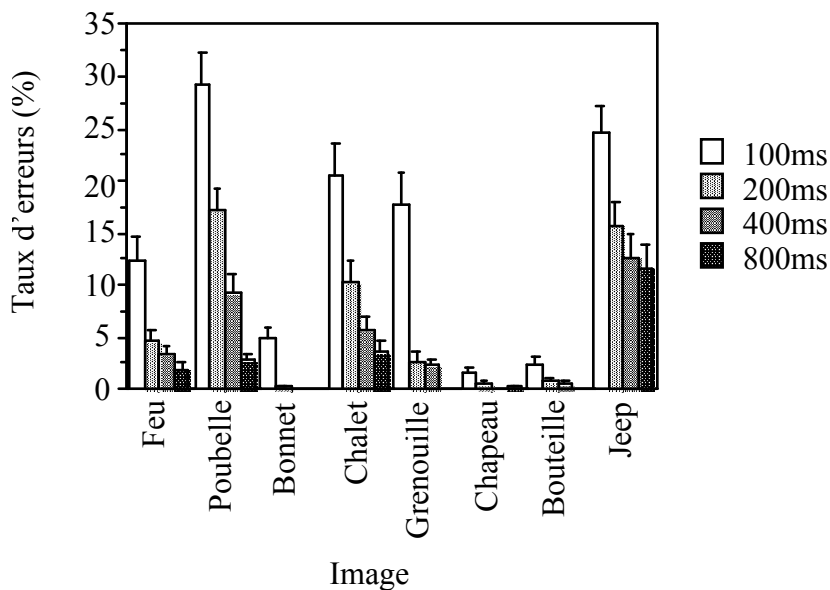


Figure 15 : Effet du temps de présentation sur le taux d'erreurs moyen obtenu avec les différentes images

4-1-3-3- Effet du contexte

Toutes conditions expérimentales confondues, les sujets font, en moyenne, plus d'erreurs lorsque l'arrangement spatial des éléments de l'image est perturbé, c'est à dire dans les images déstructurées (figure 16). Le taux d'erreurs passe ainsi de $6,02 \pm 0,42\%$ pour les images structurées à $7,54 \pm 0,47\%$ pour les images déstructurées.

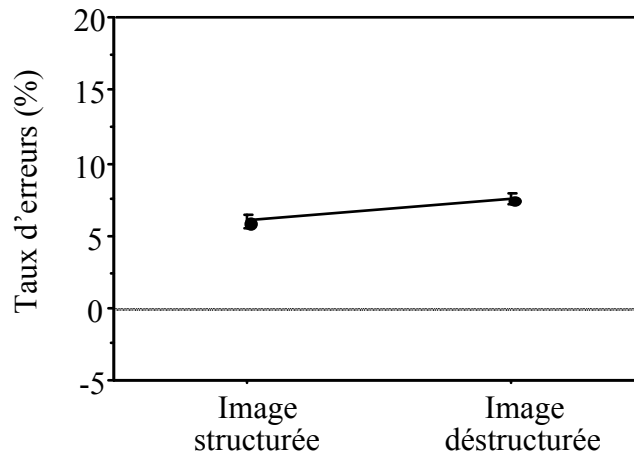


Figure 16 : Effet du contexte sur le taux d'erreurs moyen.

La différence de performance dans la tâche de reconnaissance-localisation reste modérée entre les conditions structurée et déstructurée. D'un point de vue statistique, elle est cependant significative (effet Contexte : $F^1_{2048}=23,4$; $p<0,0001$) en raison de la faible dispersion des données et de la constance de cet effet sur tous les autres facteurs de variation comme en témoignent les graphes d'interaction entre les différents facteurs des figures 17, 18 et 19.

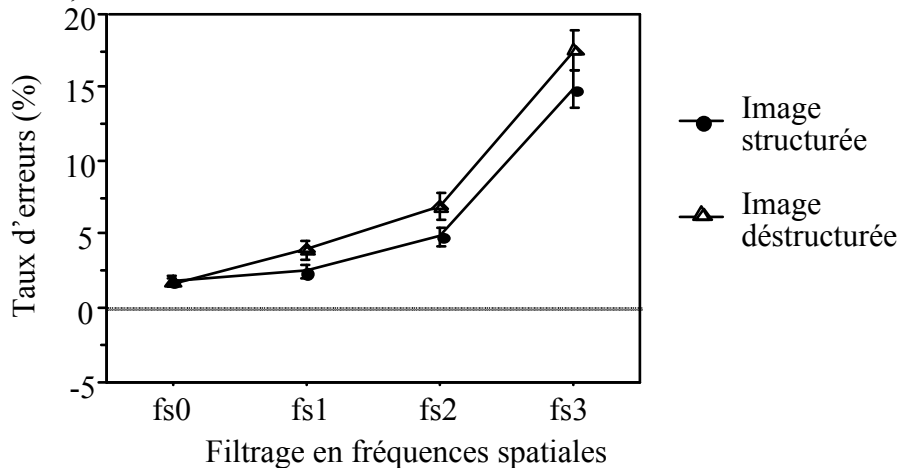


Figure 17 : Effet du filtrage en fréquences spatiales sur le taux d'erreurs moyen selon que l'image est structurée ou non.

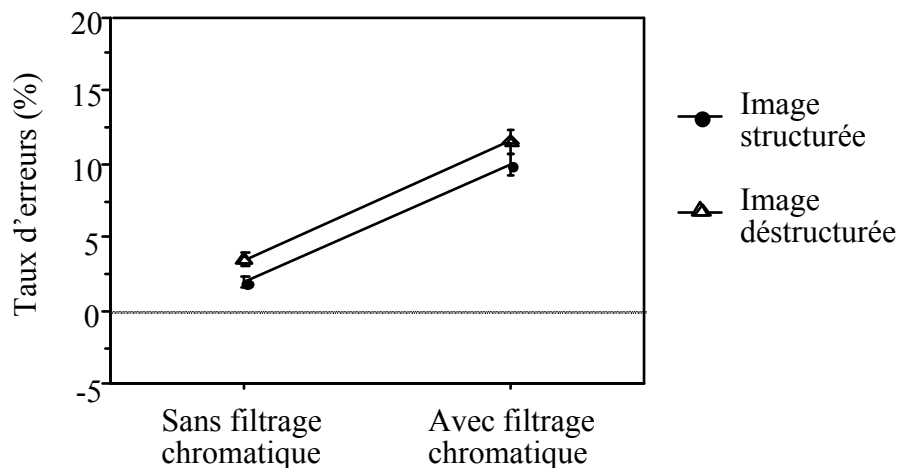


Figure 18 : Effet du filtrage chromatique sur le taux d'erreurs moyen selon que l'image est structurée ou non.

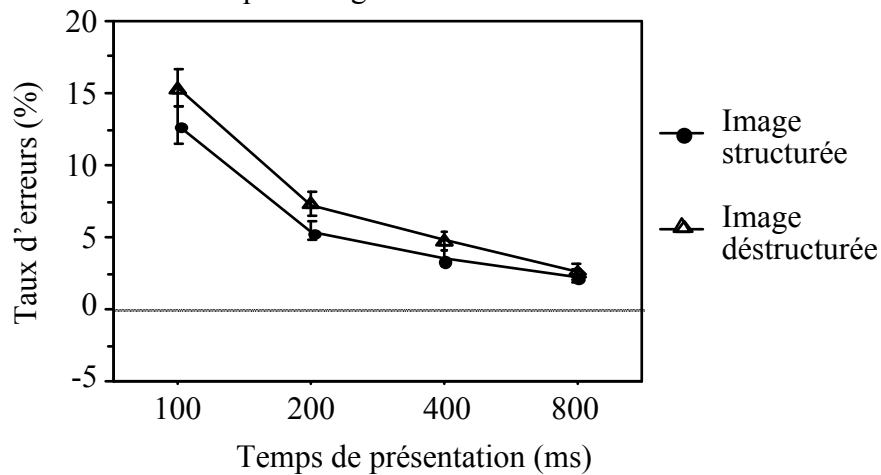


Figure 19 : Effet du temps de présentation sur le taux d'erreurs moyen selon que l'image est structurée ou non.

Au moins dans le cas du filtrage en fréquences spatiales et dans la contrainte temporelle de présentation des images, il apparaît que l'effet du contexte est d'autant plus marqué que les images sont appauvries dans leur contenu physique (niveau de filtrage plus élevé) ou que les conditions d'acquisition du signal visuel sont plus restreintes (réduction du temps de présentation).

La prise en compte des particularités liées à chaque image met en évidence que les écarts de performance entre les présentations d'images structurées et déstructurées se manifestent essentiellement au travers de 2 images : celle où les sujets doivent reconnaître et localiser la poubelle et celle où la tâche porte sur le feu tricolore. Le taux d'erreurs passe de $10,49 \pm 1,38$ à $18,75 \pm 1,79\%$ lorsque l'image est déstructurée ; la variation n'est déjà plus que de $4,65 \pm 0,93$ à $6,25 \pm 1,1$ dans le cas du feu.

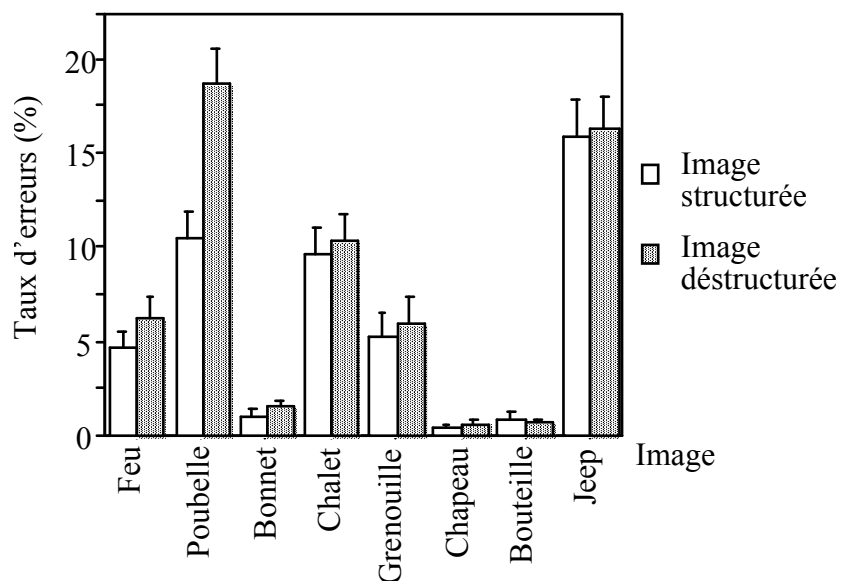


Figure 20 : Effet du contexte sur le taux d'erreurs moyen pour les différentes images.

4-1-4- Variation entre la 1^{ère} et la 3^{ème} répétition

4-1-4-1- Taux d'erreurs moyen

Les sujets ont nettement amélioré leur performance dans la tâche de reconnaissance-localisation entre la première série expérimentale et la troisième. Le taux moyen d'erreurs passe ainsi de $9,49 \pm 0,43\%$ dans la 1^{ère} répétition à $5,25 \pm 0,32\%$ dans la troisième. Une comparaison appariée des données met en évidence une différence significative du taux d'erreurs ($t_{2559} = 12,7$; $p < 0,0001$).

Cependant, cette amélioration n'est pas homogène sur l'ensemble des images : elle est d'autant plus marquée que le taux d'erreurs était initialement élevé. Deux images se détachent particulièrement, l'une contenant le feu tricolore, l'autre la poubelle, pour lesquelles le taux d'erreurs est réduit de plus de 50% (figure 21).

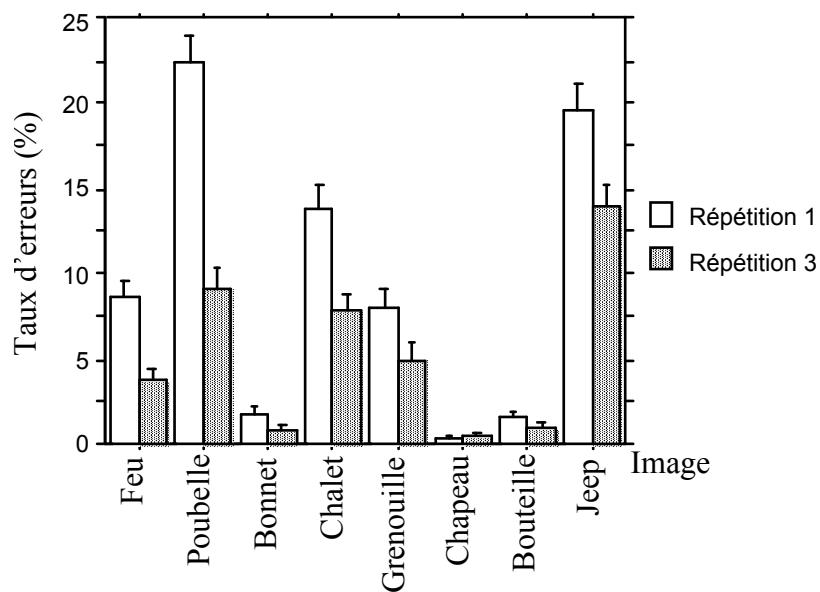


Figure 21 : Variation du taux d'erreurs moyen entre la 1^{ère} et la 3^{ème} série expérimentale.

4-1-4-2- Effet d'une altération dans la prise d'information visuelle

L'effet des différents facteurs altérant la prise d'information (filtrage en fréquences spatiales, filtrage chromatique, réduction du temps de présentation) est moins marqué sur la dégradation de la tâche de reconnaissance-localisation des divers objets dans la troisième répétition qu'il ne l'était dans la première. Néanmoins, les différences de taux d'erreurs dues à ces facteurs de variation restent statistiquement significatives (figure 22 à 24).

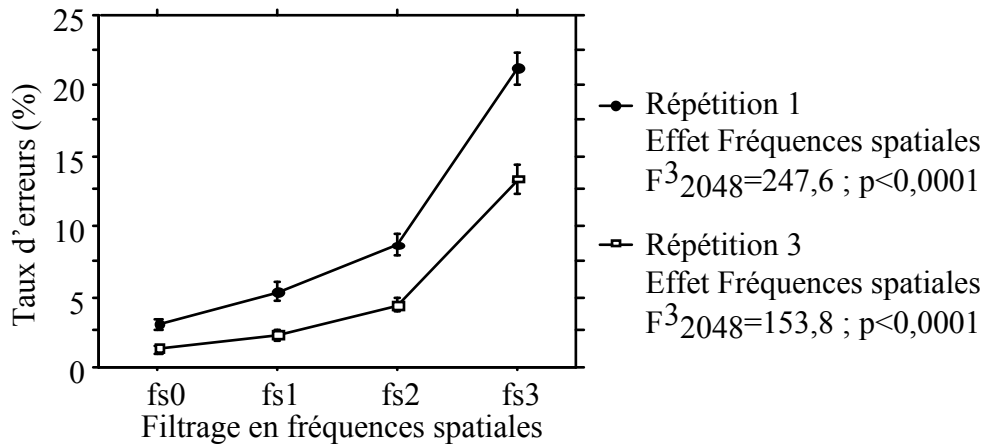


Figure 22 : Effet du filtrage en fréquences spatiales sur le taux d'erreurs moyen à la 1^{ère} et la 3^{ème} série expérimentale.

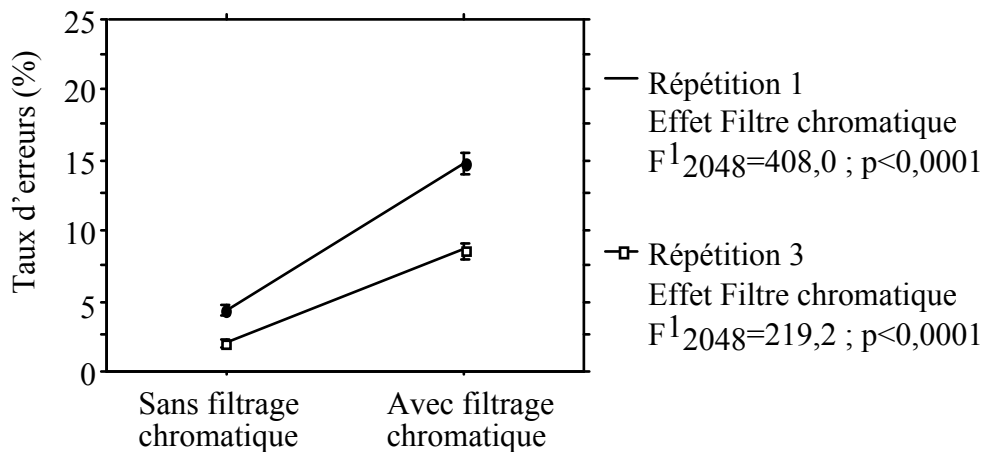


Figure 23 : Effet du filtrage chromatique sur le taux d'erreurs moyen à la 1^{ère} et la 3^{ème} série expérimentale.

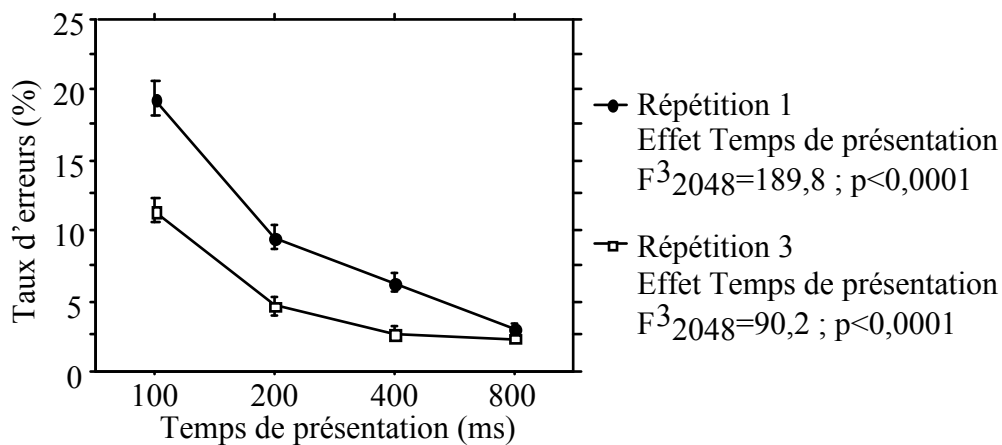


Figure 24 : Effet du temps de présentation sur le taux d'erreurs moyen à la 1^{ère} et la 3^{ème} série expérimentale.

Les sujets ont amélioré leur performance entre la première et la dernière session expérimentale ; cette amélioration est d'autant plus marquée que la tâche était

initialement détériorée (conditions de filtrage maximal, temps de présentation bref). Les performances ont donc tendance à s'uniformiser.

4-1-4-3- Effet du contexte

Là encore, le taux d'erreurs selon que l'image est présentée sous sa forme structurée ou déstructurée varie dans des proportions notablement différentes entre la première et la troisième répétition de la session expérimentale, à tel point que la différence statistiquement significative dans le premier cas, ne l'est plus dans le second (figure 25). La reconnaissance d'objets dans les images, dépendante de l'organisation cohérente des objets dans l'image en début d'expérimentation, s'affranchit du contexte en fin d'étude.

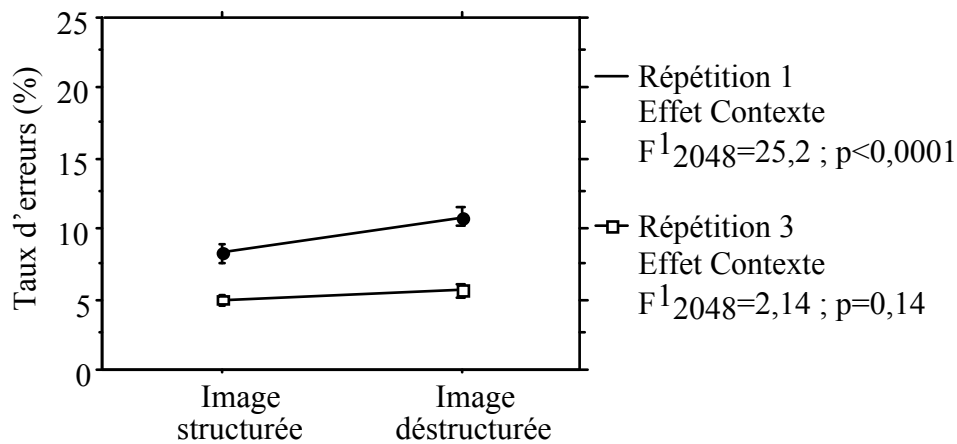


Figure 25 : Effet du contexte sur le taux d'erreurs moyen à la 1^{ère} et la 3^{ème} série expérimentale.

4-1-5- Cas particulier de certaines images

Deux images (celles contenant la jeep et le feu tricolore) se caractérisent par la présence simultanée d'un objet présentant une répartition spatiale des caractéristiques photométriques proche de celle de l'objet recherché. Pour la jeep, il s'agit du véhicule de transmission situé à la gauche de l'objet recherché ; pour le feu tricolore, c'est le fanion situé dans le coin inférieur gauche de l'image.

La répartition des erreurs a été analysée sur les présentations où l'objet recherché et le distracteur étaient simultanément affichés. Le tableau suivant présente les pourcentages de pointage de l'objet et du distracteur pour les différents traitements d'images, dans le cas de l'image de la jeep.

Tableau 2 : Taux de pointages en présence du distracteur.

		Pointage jeep	Pointage distracteur	Pointage autre
Tous traitements		80%	16,7%	3,3%
Sans filtrage chromatique	fs0	100%	0%	0%
	fs1	98,3%	2,7%	0%
	fs2	96,2%	3,8%	0%
	fs3	45%	45,4%	9,6%
Avec filtrage chromatique	fs0	95,8%	3,4%	0,8%
	fs1	89,2%	5,8%	5%
	fs2	80,4%	17,1%	2,5%
	fs3	37,5%	54,6%	7,9%

Il existe une transition nette dans la performance des sujets. Alors que les erreurs sont absentes ou en proportion négligeable jusqu'à un niveau de filtrage en fréquences spatiales centré sur 3cpd, le taux d'erreurs s'accroît brutalement lorsque l'image est amputée de son contenu en fréquences spatiales supérieures à 1,5cpd. L'ambiguïté entre l'objet recherché et le distracteur est alors totale, conduisant à une proportion de pointages identique sur les deux éléments, voire plus élevée dans le cas du distracteur (cas des images filtrées en fréquences spatiales et en contenu chromatique). Cette tendance à désigner le distracteur dans les images très dégradées peut être rapportée à un contraste de luminance plus élevé du distracteur par rapport à l'objet cible dans ce type d'images.

Les données recueillies sur les images contenant le feu tricolore sont comparables à celle de la jeep et plus contrastées encore puisque les erreurs dans la tâche de reconnaissance n'apparaissent que pour le niveau de filtrage fréquentiel maximal. Dans ces conditions, la désignation du feu et du distracteur se fait en proportions 72%-20% en l'absence de filtrage chromatique et 43,7%-42,7% en présence de filtrage chromatique. La même ambiguïté objet recherché-distracteur apparaît pour les images les plus filtrées, l'information chromatique (feu rouge) servant alors d'élément discriminant dans les images non filtrées chromatiquement.

4-2- Temps de réponse

Le temps de réponse est analysé sur l'ensemble des présentations où le pointage a désigné la cible (réponses correctes), soit 23040 présentations desquelles il faut soustraire 1568 erreurs, 84 mauvaises manipulations de l'écran tactile par les sujets, ce qui donne 21388 temps de réponses analysés. Les résultats sont exprimés en moyenne de temps de réponse et son erreur standard.

Il faut noter une grande concordance entre les résultats présentés sur le taux d'erreurs et rendant compte des échecs dans la tâche, et ceux relatifs au temps de réponse en rapport avec les essais réussis.

4-2-1- Variabilité liée aux sujets et aux images

Les cinq sujets présentent des temps de réponse qui varient en moyenne de $738,4 \pm 5,4$ à $903,1 \pm 5,2$ ms (figure 26). Cette différence de temps de réponse entre les sujets est significative (effet Sujet : $F^4_{21383}=202,7$; $p < 0,0001$). En dépit de cette variabilité inter-individuelle, la variabilité intra-individuelle reste très modérée comme en témoignent les erreurs standards associées aux valeurs moyennes. Il existe donc des sujets qui répondent rapidement, d'autre plus lentement, mais avec constance.

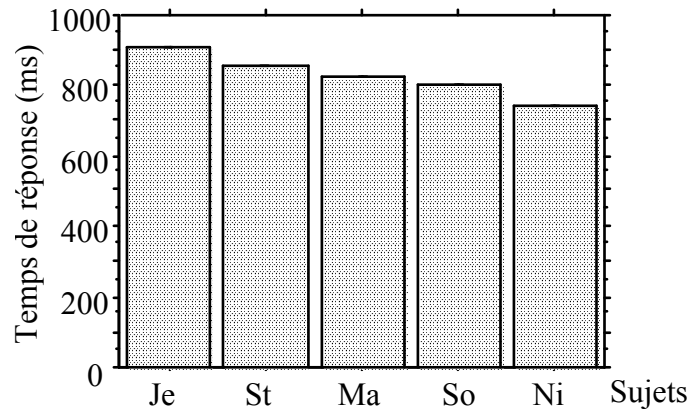


Figure 26 : Temps de réponse moyen des sujets.

De même que pour les taux d'erreurs, le temps de réponse est très variable selon l'image envisagée, allant de $766,1 \pm 3,9$ ms en moyenne pour l'image contenant le chapeau, à $893,4 \pm 7,2$ ms pour l'image contenant la poubelle (figure 27). Cette différence est statistiquement significative (effet image : $F^7_{21380}=78,0$; $p < 0,0001$). On remarque une parfaite superposition entre le taux d'erreurs associé aux images et le temps de réponse moyen, témoignant d'une plus grande difficulté dans la tâche pour certaines images comme la reconnaissance de la poubelle ou celle de la jeep.

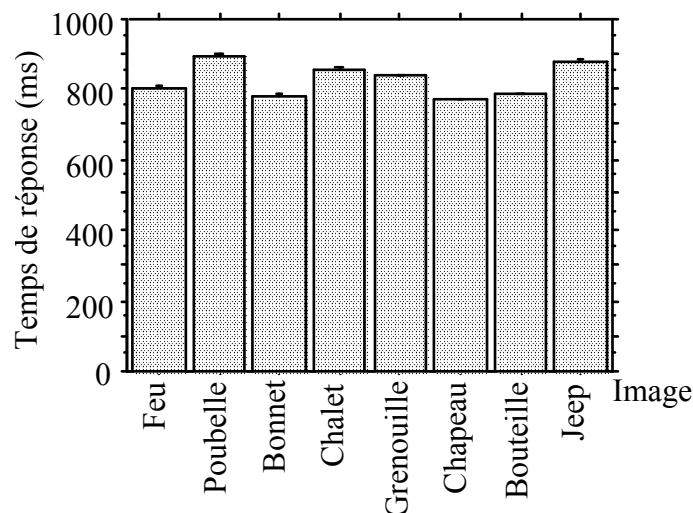


Figure 27 : Temps de réponse pour les différentes images.

4-2-2- Facteurs de variation sur l'ensemble des données

4-2-2-1- Effet de la dégradation de l'image

4-2-2-1-1- Dégradation du contenu en fréquences spatiales

Le temps de réponse moyen augmente progressivement au fur et à mesure que l'image est filtrée, passant de $792,0 \pm 3,2$ ms lorsque les images ne sont pas filtrées, à $876,4 \pm 5,0$ ms pour les images les plus filtrées (figure 28). Cet effet est statistiquement significatif (effet Fréquences spatiales : $F^3_{21384} = 88,4$; $p < 0,0001$). Comme pour le taux d'erreurs, la transition la plus marquée sur le temps de réponse se produit lors de l'application du filtrage jusqu'à 1,5cpd (fs3), avec une augmentation du temps de réponse de l'ordre de 60ms par rapport au niveau de filtrage précédent.

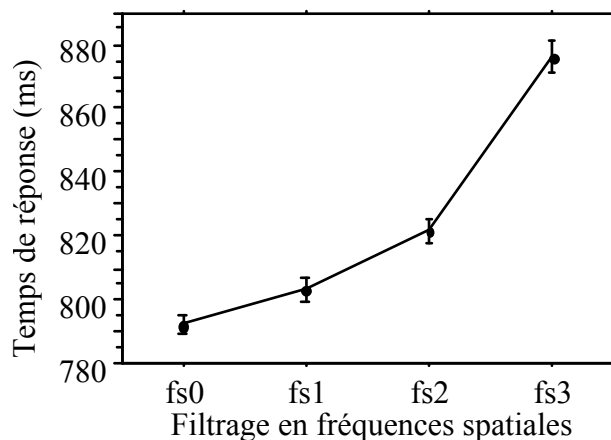


Figure 28 : Effet du filtrage en fréquences spatiales sur le temps de réponse.

4-2-2-1-2- Dégradation du contenu chromatique

Toutes conditions confondues, le temps de réponse est augmenté de l'ordre de 70ms lorsque les images sont soumises à un filtrage chromatique, sa valeur moyenne passant alors de $788,1 \pm 2,4$ à $857,7 \pm 3,1$ ms (figure 29). L'effet est significatif (effet Filtrage chromatique : $F^1_{21386} = 324,1$; $p < 0,0001$)

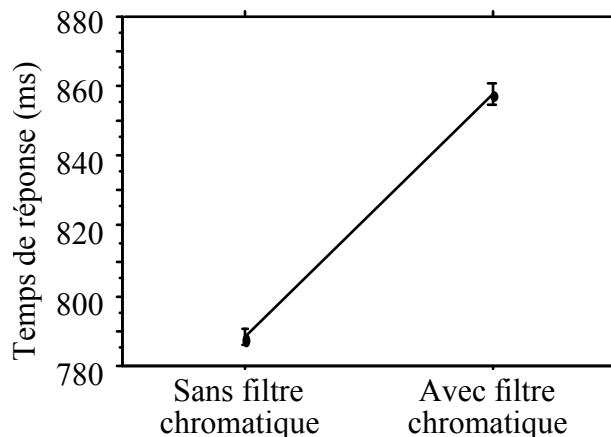


Figure 29 : Effet du filtrage chromatique sur le temps de réponse.

La dégradation du contenu physique, fréquentiel et chromatique, des images entraîne un allongement notable du temps de réponse.

4-2-2-2- Effet du temps de présentation

La performance est altérée, mais de façon moins nette, au fur et à mesure que l'on applique une contrainte temporelle sur l'affichage des images : le temps de réponse s'allonge progressivement (figure 30). De $814,0 \pm 3,7$ ms lors d'un affichage de 800ms, il atteint $832,5 \pm 4,4$ ms quand la présentation est réduite à 100ms. L'effet est encore statistiquement significatif (effet Temps de présentation : $F^3_{21384}=3,8$; $p=0,009$).

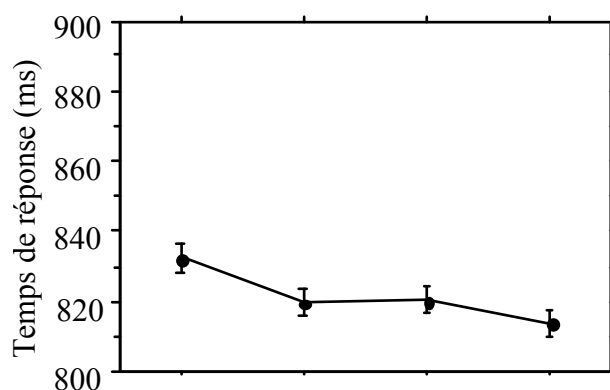


Figure 30 : Effet du temps de présentation sur le temps de réponse.

4-2-2-3- Effet du contexte

Toutes conditions expérimentales confondues, les sujets ont, en moyenne, un temps de réponse plus long lorsque l'arrangement spatial des éléments de l'image est perturbé, c'est à dire dans les images déstructurées (figure 31). Le temps de réponse passe ainsi de $816,4 \pm 2,7$ ms pour les images structurées, à $826,4 \pm 2,8$ ms pour les images déstructurées (effet Contexte : $F^1_{21386}=6,6$; $p=0,01$).

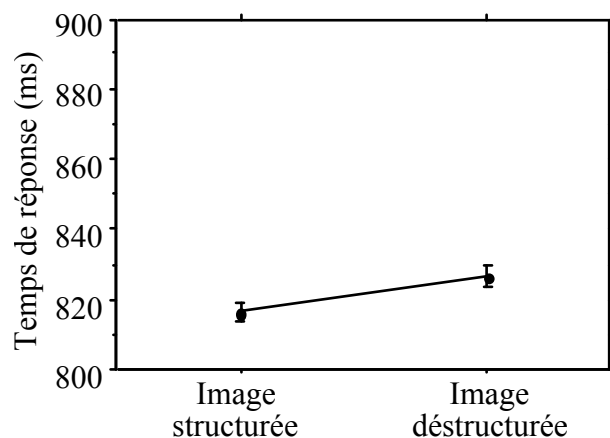


Figure 31 : Effet du contexte sur le temps de réponse.

Les 10ms en moyenne d'allongement du temps de réponse lors de la perturbation du contexte se manifestent sur les différentes altérations dans la prise d'information visuelle, et essentiellement sur les dégradations physiques de l'image, comme en témoignent les graphes d'interaction du facteur Contexte avec les autres facteurs de variation présentés dans les figures 32 à 34.

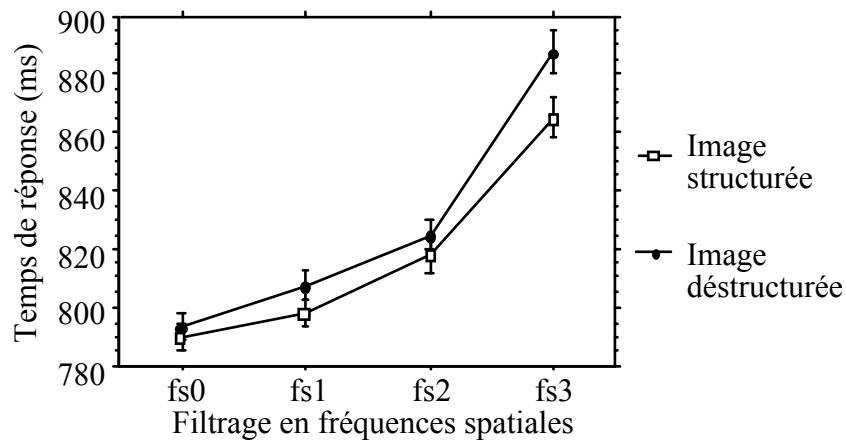


Figure 32 : Effet du filtrage en fréquences spatiales sur le temps de réponse selon que l'image est structurée ou non.

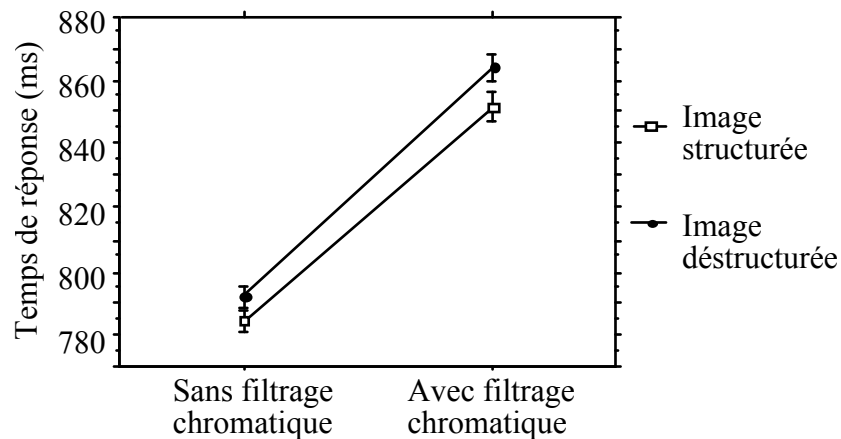


Figure 33 : Effet du filtrage chromatique sur le temps de réponse selon que l'image est structurée ou non.

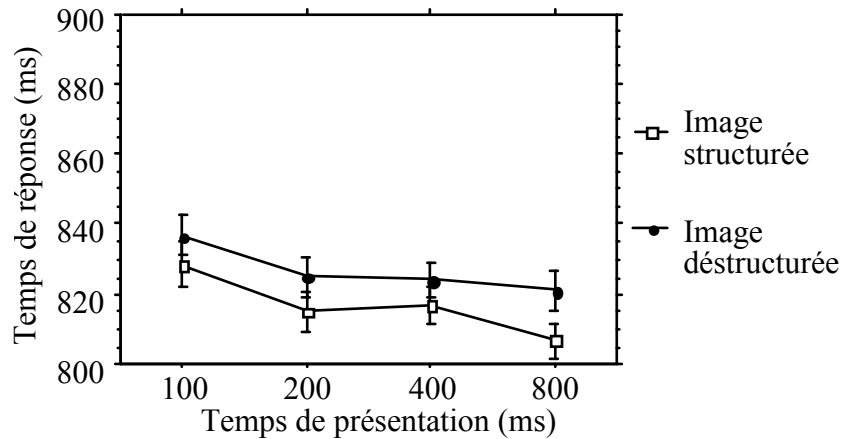


Figure 34 : Effet du temps de présentation sur le temps de réponse selon que l'image est structurée ou non.

Comme pour le taux d'erreurs, l'effet du contexte est d'autant plus marqué que les images sont appauvries dans leur contenu physique (niveau de filtrage plus élevé, ou filtrage chromatique). En revanche, aucune majoration de l'effet du contexte sur le temps de réponse lorsque le temps d'acquisition de l'information est progressivement restreint n'est observée (figure 34).

4-2-3- Variation entre la 1^{ère} et la 3^{ème} répétition

4-2-3-1- Temps de réponse moyen

On assiste à une réduction progressive du temps de réponse moyen au fur et à mesure de la répétition des séries expérimentales (figure 35). Cet effet se manifeste principalement dès la deuxième répétition (effet Répétition : $F^2_{23037}=953,4$; $p<0,0001$). Il en résulte une réduction de l'ordre de 350ms du temps de réponse en moyenne entre la 1^{ère} et la 3^{ème} répétition des séries expérimentales.

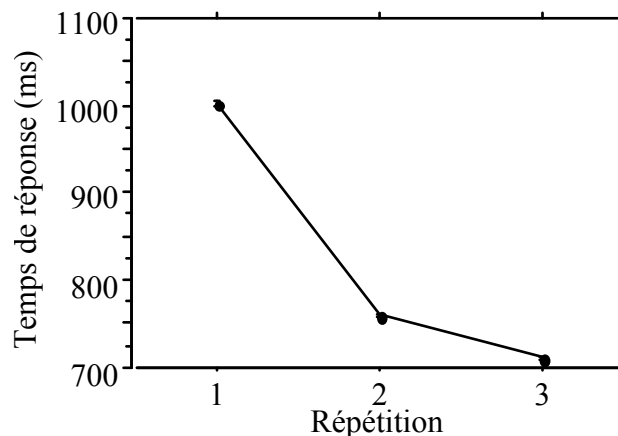


Figure 35 : Évolution du temps de réponse moyen au cours des séries expérimentales.

4-2-3-2- Effet d'une altération dans la prise d'information visuelle

L'effet des différents facteurs altérant la prise d'information (filtrage en fréquences spatiales, filtrage chromatique, réduction du temps de présentation) sur l'allongement du temps de réponse lors d'une tâche de reconnaissance-localisation des divers objets est moins marqué dans la troisième répétition qu'il ne l'était dans la première. Néanmoins, les différences de temps de réponse dues à ces facteurs de variation restent statistiquement significatives (figure 36 à 38).

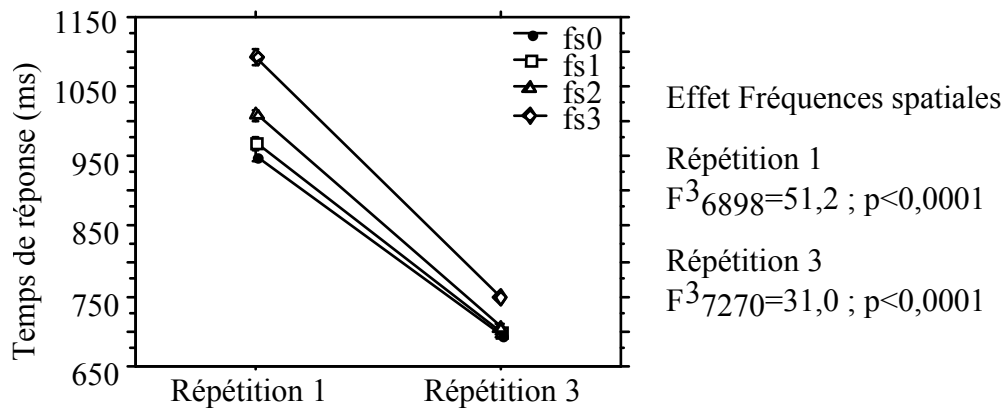


Figure 36 : Effet du filtrage en fréquences spatiales sur le temps de réponse dans la 1^{ère} et la 3^{ème} série expérimentale.

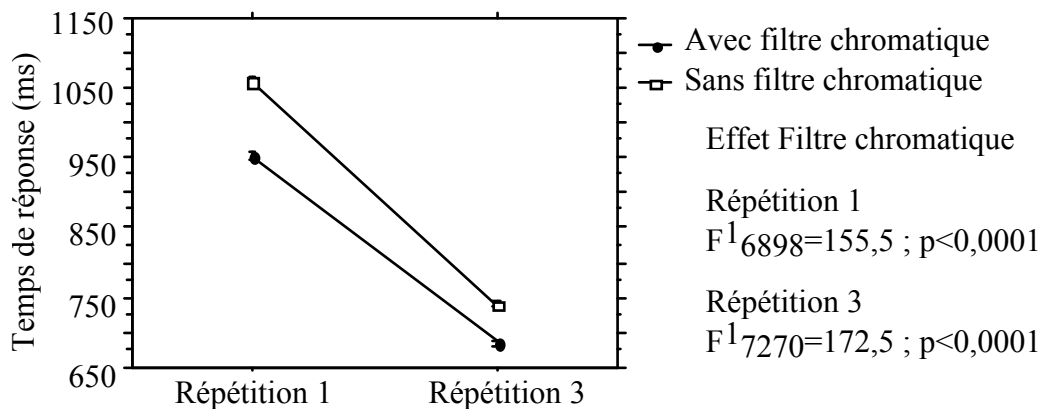


Figure 37 : Effet du filtrage chromatique sur le temps de réponse dans la 1^{ère} et 3^{ème} série expérimentale.

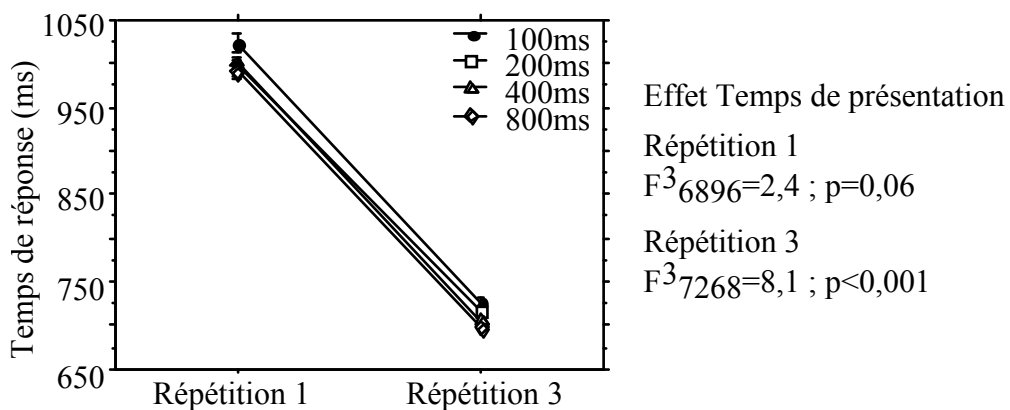


Figure 38 : Effet du temps de présentation sur le temps de réponse dans la 1^{ère} et la 3^{ème} série expérimentale.

4-2-3-3- Effet du contexte

En revanche, la présence ou l'absence d'informations sur l'arrangement spatial des différents éléments de l'espace n'influence plus le temps de réponse dans la troisième répétition (figure 39).

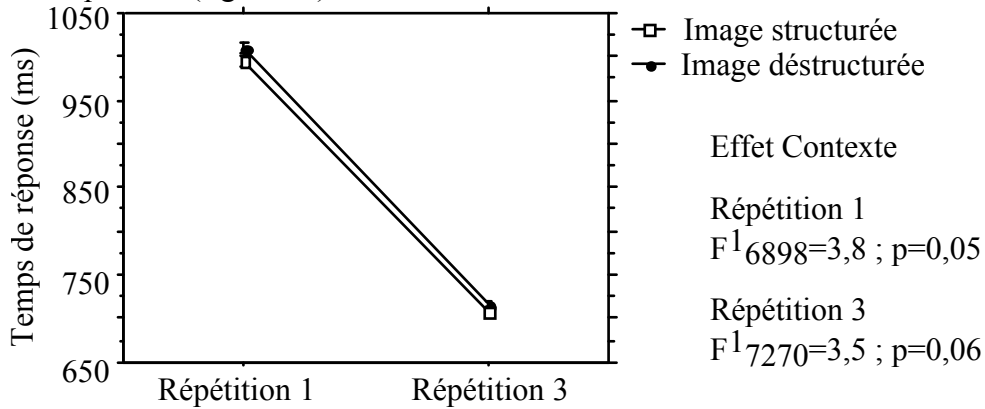


Figure 39 : Effet du contexte sur le temps de réponse à la 1^{ère} et la 3^{ème} série expérimentale.

5- DISCUSSION

Les résultats de cette étude montrent que plus on dégrade la qualité de l'information ascendante par des filtrages spatiaux et chromatiques ou par une diminution du temps d'exposition à la stimulation, plus la performance de reconnaissance est altérée. Cette altération est déterminée par l'augmentation de la proportion d'échecs de pointage (taux d'erreurs) et du temps nécessaire aux sujets pour effectuer correctement la tâche demandée (temps de réponse).

La répercussion de ces dégradations dépend de la qualité de l'information contextuelle contenue dans l'image. La proportion d'erreurs et le temps de désignation s'accroissent de façon plus importante pour les images dont l'arrangement spatial des différents éléments est perturbé, c'est-à-dire les images déstructurées.

Enfin, au cours de la répétition des différentes conditions, on constate une amélioration globale des performances de reconnaissance des sujets. Cette amélioration est associée à une diminution des différences entre les images structurées et les images déstructurées, quel que soit le niveau de dégradation. La reconnaissance d'objets dans les scènes présentées, dépendante de l'organisation spatiale cohérente des différents éléments isolables de l'image en début d'expérience, s'affranchit des informations contextuelles en fin d'étude.

5-1- Méthodologie

Les critères sélectionnés pour rendre compte de la performance en reconnaissance visuelle des sujets (taux d'erreurs et temps de réponse) réalisent une description relativement exhaustive de la performance des sujets en décrivant d'une part les conditions où la tâche ne peut pas être réalisée (taux d'erreurs) et d'autre part la difficulté à réaliser la tâche dans les cas de succès (temps de réponse). Le choix des critères d'évaluation s'est avéré judicieux. On constate en effet une bonne homogénéité des résultats obtenus sur le taux d'erreurs et le temps de réponse.

Les résultats montrent qu'il n'existe pas de liaison entre la couleur de l'objet et/ou de son fond et la difficulté de la tâche dans l'ensemble. La baisse de performance causée par le filtrage chromatique semble plus liée à la diminution de la luminance de ces images par rapport au fond d'adaptation qu'à la diminution des informations de couleur. On aurait pu faire un filtrage chromatique pur en conservant la luminance moyenne des stimulation filtrée en couleur par rapport à leur référence et évaluer plus précisément l'influence de la couleur dans une tâche de reconnaissance. Le choix n'a pas été celui-là car l'objectif était d'évaluer le rôle d'une perturbation de l'information chromatique dans des conditions naturelles et courantes. Or, les conditions naturelles qui conduisent à une altération du contenu chromatique des images (port d'optiques filtrantes) induisent indirectement une baisse de la luminance des stimulations visuelles.

Subjectivement, les sujets observant les images déstructurées étaient gênés pour reconstruire l'image présentée. Ils n'auraient pas pu identifier la sous-image sélectionnée dans la présentation. On peut être sûr que l'on perturbe au moins le contexte général de l'image. Les résultats le prouvent car il y a une influence du contexte sur la performance des sujets. Par contre, on doit être plus réservé sur l'altération du contexte local car on conserve 1/6 de la sous-image et notamment tout le fond sur lequel se détache l'objet. Des paradigmes différents, comme la localisation aléatoire de l'objet dans l'image (protocoles de violations des règles de Biederman) ou un mélange de type puzzle où les limites des pièces correspondent au plus près aux limites des objets permettraient d'analyser cet effet.

5-2- Dégradation du contenu physique de la stimulation et contexte

Le modèle d'organisation du système visuel globalement admis dans des tâches de recherche visuelle ou des tâches d'identification d'objets ou de scènes propose une collaboration (plus ou moins équitable selon les auteurs) entre les informations ascendantes issues des capteurs rétiniens et les informations descendantes issues de la mémoire à long terme ou des processus attentionnels. Les informations sensorielles constituant la source de toute la perception, on peut suggérer qu'une détérioration de la qualité de ces informations se traduit par une perturbation globale des processus de traitement et donc par une baisse des performances. Les résultats exposés dans cette étude confirment effectivement cette hypothèse. La dégradation du contenu physique de la stimulation, tant au niveau du spectre fréquentiel qu'au niveau des contrastes de luminance et de chrominance, affecte négativement les processus de reconnaissance en

les rendant moins fiables (augmentation du taux d'erreurs) et plus longs (élévation du temps de réponse). Les résultats montrent également qu'une contrainte temporelle réalisée par une exposition au stimulus de plus en plus courte se traduit par une baisse des performances. De plus, on assiste à une potentialisation des effets : l'accumulation des paramètres pénalisant la prise d'information rend la tâche de reconnaissance-localisation demandée au sujet de plus en plus difficile.

Très peu d'études se sont intéressées à la reconnaissance ou à l'identification de scènes naturelles dégradées. Les seuls résultats comparables que l'on trouve dans la littérature portent certes sur des tâches de haut niveau mais pour des stimulations plus simples telles que des lettres ou des objets isolés et très rarement des images naturelles de visages constituant un sous-groupe très particulier.

Concernant le contenu en fréquences spatiales des images, les résultats sont contradictoires. Pour certains, les basses fréquences spatiales sont suffisantes pour reconnaître des lettres ou des visages et que la présence ou l'absence des informations de détail contenues dans les hautes fréquences n'influence pas les performances du fait de leur redondance (Ginsburg, 1978). Au contraire, Fiorentini, Maffei et Sandini (1983) soutiennent que l'addition des hautes fréquences favorisent la tâche et de plus qu'elles sont suffisantes pour reconnaître des visages.

Dans la présente étude, les résultats montrent qu'une amputation des informations dès les moyennes fréquences spatiales (de l'ordre de 1,5cpd) perturbe significativement les performances des sujets. Les données recueillies ne permettent pas de déterminer le rôle des différentes fréquences spatiales pour la reconnaissance d'objets dans des scènes naturelles. Elles permettent seulement de conclure que des informations utiles à la reconnaissance visuelle d'objets sont présentes et disponibles dans les hautes fréquences spatiales (au-delà de 1,5cpd).

Si on se réfère au schéma général de l'organisation du système visuel dans la perception des formes, les informations ascendantes et descendantes collaborent à l'élaboration d'une représentation interprétable du stimulus. On peut donc suggérer l'hypothèse que la baisse du poids informatif de l'une des deux voies soit compensée par une augmentation d'importance de la voie non affectée.

Ainsi peut-on supposer d'une part qu'un stimulus vu pour la première fois et ne disposant pas encore de représentation mémorisée doit nécessiter un temps de présentation plus long qu'un stimulus connu pour accéder à la structuration perceptive du signal sensoriel, utile à son incorporation au domaine des connaissances. D'autre part, on peut suggérer que la baisse de qualité de l'information sensorielle augmente le poids de la voie descendante, et notamment des connaissances sur l'organisation spatiale des éléments d'une scène naturelle, afin de pallier le manque d'information sensorielle et de pouvoir effectuer la tâche. Les résultats de cette expérience sont en faveur de cette deuxième hypothèse.

En effet, on constate, non seulement pour les taux d'erreurs mais aussi pour les temps de réponse, qu'une dégradation du contenu physique des images (filtrage en fréquences spatiales, filtrage chromatique) affecte d'une façon significativement plus importante les images déstructurées que les images structurées. On en conclut que lorsque la voie ascendante n'apporte pas suffisamment d'informations, il y a renforcement du poids de la voie descendante et recours majoré aux informations de contexte.

Certains auteurs suggèrent ce mode de fonctionnement pour des tâches de lecture. Une hypothèse proposée par Stanovich (1980), sur la base d'observations faites chez de jeunes lecteurs plus ou moins habiles, suppose l'existence d'un "processus interactif compensatoire" permettant de rendre compte des effets facilitateurs du contexte. Ce processus descendant, s'appuyant sur d'autres sources de connaissance (notamment les informations contextuelles), prendrait plus d'importance quand les processus ascendants traitant l'information sensorielle se montrent inefficaces (Dubois & Sprenger-Charolles, 1988).

Certaines données de la littérature viennent également étayer la première hypothèse selon laquelle un manque ou une perturbation des informations de la voie descendante peuvent être compensés par une durée de fixation supérieure afin de stabiliser le stade de structuration perceptive. Biederman a en effet montré que la violation des règles sémantiques ou syntaxiques d'organisation d'une scène naturelle était difficile à trouver dans des présentations brèves (Biederman & col., 1982), mais qu'une fois cette incongruité trouvée, les sujets passaient toujours plus de temps à examiner cette partie non attendue de la scène (Loftus et Mackworth, 1978).

Les résultats de l'expérimentation concernant l'influence du contexte sur les performances des sujets montrent que cet effet ne se manifeste que dans les conditions d'un signal sensoriel afférent dégradé. Ce résultat est en accord avec l'hypothèse selon laquelle un sujet normal, dans des conditions de perception optimales (du point de vue de la qualité du contenu physique de la stimulation) peut se passer des informations de contexte pour effectuer une tâche d'identification (Thorpe, 1995). Dans le cas présent, le résultat peut être étendu au cas d'une tâche de reconnaissance-localisation dans une image naturelle.

Il faut néanmoins nuancer cette conclusion, car si on peut justifier expérimentalement l'importance croissante des informations de contexte dans le cas d'une perception dégradée, on ne peut prouver que le système visuel n'utilise pas ces informations dans le cas des images normales. On peut simplement dire que dans les conditions où l'information sensorielle est de bonne qualité, l'information contextuelle est moins pertinente et non déterminante pour la performance dans la tâche.

Selon Oliva et Schyns (1994), les informations de détail, véhiculées par les hautes fréquences, concerneraient les informations d'identité alors que les informations plus grossières, contenues dans les basses fréquences spatiales, concerneraient les relations spatiales entre les différents éléments de la scène (ou contexte). La théorie du traitement "Coarse-to-Fine" utilisée dans leur étude propose une organisation dissymétrique entre les informations grossières et fines en supposant que les premières, traitées à un niveau plus précoce, contraignent le traitement plus tardif des secondes. Si l'on considère, comme il a été dit précédemment, que les hautes fréquences spatiales contiennent des informations utiles pour la reconnaissance, le sujet ne dispose plus, dans le cas d'images dégradées, que d'informations grossières concernant l'organisation spatiale de la scène. Sa performance doit alors être particulièrement affectée par la déstructuration de cette organisation. C'est précisément ce que l'on observe dans la présente étude où la performance est d'autant plus affectée par la perturbation du contexte que les images sont dégradées dans leur contenu physique. Les résultats de cette étude sont en accord avec la théorie "Coarse-to-Fine" du traitement de l'information visuelle dans une tâche de reconnaissance-localisation.

En ce qui concerne le temps de présentation des images, les données de psychophysique sur la recherche visuelle et les données de neurophysiologie suggèrent que des temps d'exposition longs favorisent le traitement des hautes fréquences spatiales tandis que des temps d'exposition courts favorisent le traitement des basses fréquences spatiales (Paquet & Merikle, 1984). On suppose donc, dans le cadre de cette expérience, que pour le temps de présentation le plus court (100ms), le traitement visuel est préférentiellement effectué sur les basses fréquences spatiales et que l'influence du contexte y est plus marquée qu'à 800ms. Les résultats obtenus ne vont pas dans ce sens ce qui n'est pas contradictoire avec les données de la littérature. En effet, Oliva et Schyns (1994) ont montré que des temps de présentations inférieurs ou égaux à 30ms étaient nécessaires pour favoriser spécifiquement les basses fréquences spatiales. Or, tous les temps de présentation de l'étude se situent au-delà de cette limite.

Les résultats de cette expérience permettent de conclure que plus on dégrade la prise d'information en altérant le contenu fréquentiel ou chromatique de la stimulation, ou plus on contraint la prise d'information en réduisant le temps d'exposition, plus on perturbe le processus de reconnaissance-localisation de l'objet-cible. Dans ces conditions d'information ascendante appauvrie, les informations de haut niveau concernant le contexte de la scène (organisation spatiale des différents éléments isolables) suppléent une partie de ce manque. En revanche, dans le cas d'une perception normale, le poids des données contextuelles s'avère négligeable pour rendre compte de la performance visuelle.

5-3- Apprentissage et familiarité

L'adaptabilité est l'une des caractéristiques principales du système visuel. Cette adaptabilité se traduit dans les résultats par une amélioration importante des performances (taux d'erreurs et temps de réponse) au fil des présentations. Il y a donc un effet d'apprentissage de la tâche.

Une mention toute spéciale doit être faite concernant l'influence du contexte sur la reconnaissance des objets dans des conditions de signal sensoriel dégradé. On remarque que si le contexte influence de façon significative le taux d'erreurs et les temps de réponse dans la première répétition de la session expérimentale, ce n'est plus le cas dans la troisième répétition. Les stratégies de traitement de l'information semblent donc avoir changé au fur et à mesure que les sujets devenaient de plus en plus familiers avec les stimulations utilisées et de plus en plus experts dans la tâche de reconnaissance-localisation. Cette évolution des résultats peut s'expliquer dans le cadre général du modèle de la perception des formes, en supposant l'élaboration, par les sujets, de "configurations de primitives" constituées chacune de l'ensemble des caractéristiques psychophysiques discriminantes de l'objet recherché. La configuration de primitives spécifique de l'objet permet ainsi de le distinguer de son environnement malgré des dégradations du contenu physique de la scène. Cette supposition peut se justifier par un certain nombre de constatations expérimentales et théoriques sur la perception visuelle.

Tout d'abord, d'un point de vue théorique, des associations d'opérateurs neuronaux ont pu être mises en évidence, en neurophysiologie, qui permettent au système visuel de reconnaître un objet indépendamment de son aspect. Ces associations regroupent :

- des opérateurs neuronaux spécialisés qui extraient les informations utiles des images présentées,
- et des opérateurs neuronaux adaptables qui peuvent se spécialiser par apprentissage en apprenant, par exemple, la bonne combinaison d'opérateurs spécialisés qui donne la réponse optimale pour un type d'objet particulier en fonction du contexte de présentation (Burnod & Otto, 1994).

Dans un certain nombre d'études, la familiarité est associée à la notion de configuration et à la notion d'automatisation des processus de traitement. L'association entre familiarité et configuration est notamment très marquée dans les études sur la perception des visages. La perception des visages a fait l'objet de nombreuses recherches expérimentales, en raison du présupposé que ces stimulus constitueraient une classe particulière faisant l'objet de traitements spécifiques. Un certain nombre de constatations montrent que les visages sont traités de façon globale en tant que configurations homogènes et que la modification du contexte de perception (mélange des traits) de ces configurations détériore considérablement la performance, en obligeant le sujet à revenir à un traitement des informations plus locales. Cette interaction entre les traitements des locaux et globaux de l'information faciale est parfaitement illustrée par l'illusion dite de Thatcher (Thomson, 1980).

Dans le cadre de la présente étude, chaque sujet, pour chaque type de dégradation, élaborerait une configuration de primitives adaptée à l'objet recherché, ceci au fur et à mesure de l'exposition aux différentes conditions. On peut aussi envisager que la rapidité d'élaboration de ces configurations soit fonction non seulement de la fréquence et de la durée d'exposition à la dégradation mais aussi de la capacité d'observation et d'analyse des sujets, de la nature et de l'intensité de la dégradation ou encore de la ressemblance perceptive entre l'objet-cible et un autre élément de la scène, que l'on considérera alors comme un distracteur.

Certains résultats de l'expérience permettent d'argumenter en faveur d'une évolution progressive de la constitution des différentes configurations de primitives. Les données les plus marquantes concernent l'évolution du poids relatif des différentes fréquences spatiales sur le taux d'erreurs et sur le temps de réponse en fonction des trois répétitions. On constate qu'il y a un net resserrement des performances issues des diverses conditions de filtrage entre la première et la troisième répétition. La performance des sujets s'améliore préférentiellement dans les conditions les plus dégradées au fur et à mesure des présentations. Un effet comparable est observé pour les autres facteurs de dégradation de la prise d'information (filtrage chromatique et contrainte temporelle).

Les sujets, sélectionnés sur des critères ophtalmologiques de normalité, ont une bonne fonction visuelle et disposent, au début de l'étude, d'une configuration de primitives pour chaque objet basée sur une description pléthorique du stimulus. Cette configuration de primitives est adaptée aux capacités d'analyse spatio-temporelle des sujets mais inadaptée aux images dégradées, d'où une altération marquée de la performance. Les données expérimentales ne permettent pas de trancher entre :

- l'élaboration progressive par les sujets de configurations de primitives adaptées à chaque type de dégradation dans la prise d'information visuelle, assurant la discrimination de l'objet cible par rapport aux autres éléments de la scène ;

- ou la transformation de la configuration de primitives initiale, tenant compte des différents aspects de l'objet cible dans les diverses conditions dégradées et comportant des primitives invariantes quelles que soient les conditions de présentation de l'image.

Cette alternative pourrait être levée par une étude complémentaire analysant les possibilités de transfert d'une configuration de primitives d'une condition de dégradation à l'autre.

L'amélioration des performances au niveau du taux d'erreurs est cohérente avec l'élaboration de ces configurations de primitives pertinentes sous l'effet de l'exposition répétée. La prise en compte des résultats obtenus avec les différentes images permet de mettre en évidence certaines propriétés et limites d'efficacité des configurations de primitives.

Tout d'abord, le niveau de performance variable entre les images peut être rattaché à la difficulté d'élaboration de la configuration de primitives. L'exemple de l'image contenant la poubelle est typique à cet égard. La poubelle est, en échelle, en contraste et en couleur, très comparables aux autres objets constituant la scène. La configuration de primitives associée à la poubelle est de fait faiblement discriminante vis-à-vis des autres éléments de l'image. Il en résulte un taux d'erreurs et un temps de réponse plus élevés pour cet objet cible, traduisant directement la difficulté dans la tâche de reconnaissance-localisation. On note également une altération régulière de la performance au fur et à mesure que l'image est filtrée sans effet spécifique d'un niveau de filtrage en raison du maintien de la similarité du signal par rapport à l'ensemble de l'image sur tout le spectre fréquentiel.

Dans le cas des images contenant le feu et la Jeep, il existe un objet distracteur dans l'image (le fanion et le véhicule de transmission). L'ambiguïté cible-distracteur ne se manifeste pas dans les images initiales, ni même dans les conditions de faible dégradation des images car la configuration de primitives de la cible reste discriminante par rapport au distracteur. Par contre, dans les deux cas, l'application du troisième niveau de filtrage, enlevant les fréquences spatiales supérieures à 1,5cpd, fait disparaître les primitives discriminantes. L'ambiguïté est complète, conduisant à une répartition équiprobable de la désignation de la cible et du distracteur, voire à une désignation préférentielle du distracteur lorsque celui-ci offre un signal plus contrasté dans l'image filtrée.

Les données montrent clairement que l'apprentissage améliore les performances des sujets tant en terme de taux d'erreurs que de temps de réponse. De plus, l'accumulation des présentations, donc l'augmentation des compétences des sujets, se répercute sur le poids des informations contextuelles dans la tâche de reconnaissance-localisation des objets.

En effet, dans la troisième répétition et quel que soit le niveau de filtrage ou le temps de présentation considéré, le contexte ne joue plus aucun rôle au niveau des taux d'erreurs et des temps de réponse.

Un résultat comparable est obtenu pour la perception des visages, puisqu'il est montré que plus un visage est familier, moins l'effet du contexte intervient (Tiberghien, 1988).

Cet effet est également retrouvé dans un autre domaine, celui de la lecture chez des sujets jeunes. Ces études montrent que l'effet du contexte est fonction de la compétence du sujet et s'estompe avec l'apprentissage (Dubois & Sprenger-Charolles, 1994). Ces auteurs citent les travaux de Fodor (1986) montrant l'absence de rôle immédiat et habituel du contexte sur l'identification des mots.

Même si cette dernière remarque reste controversée dans le domaine de la lecture, l'intervention modulée du contexte dans les processus perceptifs de reconnaissance-localisation apparaît validée par les présents résultats expérimentaux. Il semble donc que l'on ait un effet variable du contexte : il intervient quand les conditions de dégradation sont telles que les informations ascendantes ne sont plus suffisantes pour permettre d'effectuer correctement la tâche demandée, il n'est plus déterminant dès que l'élaboration et la mise en œuvre des configurations élémentaires sont achevées et permet de distinguer l'objet de son environnement quelles que soient les conditions de dégradation de l'image.

Cette variabilité de l'effet Contexte affaiblit les modèles de type "Coarse-to-Fine" et ceux s'appuyant aussi sur une prédominance immuable des informations globales contenues dans les basses fréquences (Navon, 1983 ; Oliva & Schyns, 1994) qui supposent une influence du contexte résistante à l'effet d'apprentissage.

En tenant compte des spécificités liées à l'effet de l'apprentissage, on peut considérer que le modèle général de l'organisation des différentes voies visuelles dans une tâche d'identification est reproductible pour une tâche de reconnaissance-localisation d'un objet. L'approche adoptée permet d'intégrer la diversité des images naturelles et donc la difficulté variable à élaborer des configurations élémentaires caractéristiques de l'élément recherché. Ces configurations permettent d'avoir une reconnaissance résistante à des conditions perceptuelles dégradées. En fonction de la saillance de l'élément cible par rapport à son fond et en de la présence ou non d'un ou plusieurs distracteurs, les configurations sont plus ou moins faciles à établir et plus ou moins efficaces. Le choix d'images naturelles offre une grande diversité de sollicitation des conditions perceptives et une variabilité du poids des informations contextuelles et des différents facteurs d'altération des informations sensorielles.

5-4- Applications défense

Les constatations faites dans le cadre de l'effet de l'apprentissage peuvent s'appliquer au domaine de l'entraînement. Les pilotes sont sélectionnés selon des critères ophtalmologiques très strictes. Dans leurs tâches perceptives quotidiennes, ils disposent d'une excellente acuité visuelle, liée à une image rétinienne parfaitement définie. Les mécanismes de perception des formes sont donc adaptés à un flux informationnel riche en détails. Si on veut atteindre un niveau élevé de performance dans la reconnaissance d'éléments en conditions visuelles dégradées, il est préférable de ne pas limiter la tâche d'entraînement aux conditions de perception optimale. Le pilote ne peut en effet élaborer une (ou des) configuration(s) de primitives pertinente(s) pour une perception adaptée à

des conditions de prise d'information plus ou moins optimales qu'en multipliant les expositions à différents niveaux de dégradations. Il semble donc important de ne pas limiter le port des systèmes optiques aux conditions opérationnelles et proposer un entraînement spécifique simulant les contraintes liées aux différents types de visuels utilisables.

La familiarisation par entraînement, utile dans l'aérospatiale, trouve aussi certaines applications dans le domaine industriel. La configuration de primitives adaptée à l'objet évalué tient lieu de référence perceptive qui, plaquée sur le signal entrant, permet de détecter très rapidement les variations par rapport au pattern d'origine. La précision de ces configurations de primitives est le garant du bon rendement des personnes effectuant le contrôle qualité des produits finis.

Les résultats obtenus par ... confirment expérimentalement qu'un élément nouveau introduit dans un environnement familier est détecté très rapidement grâce à un traitement en parallèle des caractéristiques primitives de la scène.

-confirmation également de ce fonctionnement au niveau du développement de l'enfant
-permet l'automatisation des processus de traitement + éco car le système n'a plus à traiter l'ensemble de la scène mais seulement les éléments différents

Application médecine du travail mais aussi photointerprètes

La détection rapide de la variation par rapport à la norme (ou par rapport à l'habitude) a des enjeux importants au niveau de la médecine du travail. Ces professionnels sont entraînés à évaluer la normalité, toute modification de ce pattern de référence est immédiatement détectée puisque tenant lieu d'élément non-familier sur un fond familier.

6- CONCLUSIONS ET PERSPECTIVES

L'expérience décrite dans ce document montre que plus on pénalise la prise d'information en altérant le contenu fréquentiel ou chromatique de la stimulation ou encore en augmentant la contrainte temporelle d'exposition, plus on perturbe le processus de reconnaissance-localisation de l'objet-cible. Dans cette situation, où les informations issues des afférences sensorielles sont dégradées, les informations de contexte, c'est-à-dire les informations concernant l'organisation spatiale de la scène, vont intervenir de façon plus importante et suppléer ce manque de données ascendantes. Néanmoins, dans les conditions où le contenu physique de la stimulation est normal ou peu dégradé, les informations contextuelles n'ont pas une influence significative sur la performance.

La comparaison des données obtenues lors de la première et de la troisième répétition des présentations met en évidence que les sujets ont des stratégies de perception évolutives du fait de l'apprentissage. Les nouvelles stratégies se traduisent par l'élaboration pour chaque objet d'une "configuration de primitives" qui au fur et à mesure des expériences perceptives à différents niveaux de filtrage tient lieu d'invariant perceptif et assure la reconnaissance de l'objet en dépit des dégradations. Ces

"configurations de primitives", associées à une automatisation des processus de traitement provoquent non seulement une baisse importante des taux d'erreurs et du temps de réponse, mais aussi la disparition de l'influence du contexte sur la performance, et ceci quelles que soient les dégradations considérées.

+Petit mot sur l'entraînement spécifique=>images dégradées

6-1- Perspectives à court terme

La notion de "configuration de primitives" adaptée à la tâche de reconnaissance visuelle pourrait être précisée par une expérimentation complémentaire.

- L'étude permettrait de déterminer expérimentalement l'existence de différentes configurations ou au contraire l'élaboration d'une configuration de primitives basée sur les invariants de la cible. En familiarisant le sujet à un certain niveau de dégradation de l'image, il s'agirait de vérifier dans un second temps le maintien de la performance pour une condition moins dégradée mais perturbant la performance d'une population témoin.

- Le second aspect concerne l'évaluation de l'efficacité de la configuration de primitives en affranchissant le sujet du "priming" sémantique dans sa tâche de reconnaissance d'objet. L'annonce au sujet de la scène affichée peut en effet servir de préactivation pour certains récepteurs spécifiques des dimensions caractéristiques de l'objet présenté. Cette préactivation serait déterminante pour la performance du sujet tant que la configuration de primitives serait dans sa phase d'élaboration et non encore adaptée aux conditions d'images dégradées. Par contre, comme les informations contextuelles dans la présente étude, la préactivation ne serait plus déterminante au-delà de la phase d'apprentissage.

Cette préactivation sensorielle est à rapprocher de la séquence prémotrice associée au mouvement simulé, non exécuté où l'on assiste à une élévation dans les aires corticales motrices des populations de neurones destinées à la commande du mouvement considéré.

Dans cette deuxième orientation possible, une pré-étude a été effectuée sur un sujet expert dans la tâche de reconnaissance-localisation d'objets dans des scènes naturelles, ce sujet ayant passé l'intégralité de l'expérimentation principale. Il s'est avéré qu'en enlevant le nom signalant au sujet l'objet affiché à la présentation suivante, les performances restent inchangées en terme de taux d'erreurs. On n'assiste qu'à une légère élévation du temps de réponse.

Il apparaît nécessaire d'effectuer une étude comparative selon que le sujet dispose ou non d'un "priming" sémantique, chez des sujets experts et des sujets novices. Cette approche permet de vérifier la part de ces informations d'ordre contextuel dans une tâche de reconnaissance-localisation d'objets dans des scènes naturelles.

La comparaison des taux d'erreurs entre les images a montré une grande variabilité de difficulté perceptive, avec un maximum de difficulté pour deux images : celles qui contiennent la Jeep et la poubelle. L'expérience complémentaire mettrait préférentiellement en œuvre de telles images pour affiner l'analyse des résultats.

6-2- Perspectives à long terme

Au delà de l'influence des fréquences spatiales sur les seuils de perception des contrastes, on peut constater que le contenu fréquentiel détermine de façon importante la qualité de reconnaissance des objets dans une scène naturelle. Si certaines études ont pu mettre en évidence la forte influence des basses fréquences spatiales, véhiculant les informations d'organisation spatiale de la scène, dans une tâche d'identification, on a montré ici que cette dominance pouvait être contestée et qu'il y avait une part non négligeable des hautes fréquences destinées à cette tâche de reconnaissance. Il se peut néanmoins qu'en fonction de fortes contraintes environnementales (temps d'exposition très courts, éblouissement, réduction du champ visuel, ...) ce poids relatif entre les basses et hautes fréquences spatiales soit modifié du fait d'une modification de la sensibilité du système visuel. De même que les processus de traitement évoluent au cours de l'entraînement à une reconnaissance d'objets dans des images dégradées, ils peuvent s'adapter aux variations de sensibilité associées aux conditions de stimulation.

On peut aussi noter que, mise à part l'étude de Praje, Tjan et Legge (1995), l'ensemble des travaux porte sur les poids relatifs des hautes et des basses fréquences spatiales dans différentes tâches de perception visuelle. On peut s'étonner que les moyennes fréquences, celles pour lesquelles le système visuel présente un maximum de sensibilité (Fonction de Sensibilité aux Contrastes), n'apparaissent pas dans la répartition des rôles.

Les travaux seront donc poursuivis pour déterminer les poids respectifs de ces trois canaux fréquentiels dans une tâche contrôlée de reconnaissance d'objets dans des scènes visuelles naturelles, et estimer l'éventuelle variation des poids sous l'effet de fortes contraintes environnementales.

La perception visuelle des images naturelles constitue la motivation de l'ensemble des travaux menés sur la perception visuelle des formes. Cependant, des images frustrées ou élémentaires leur sont généralement préférées dans le but de contrôler plus étroitement certains paramètres expérimentaux. L'étude menée dans le cadre du DEA de Sciences Cognitives, a mis en évidence que les modèles généraux développés pour rendre compte de la perception des formes sur des images élémentaires gardaient leur validité dans une tâche de reconnaissance d'objets dans des images naturelles. L'altération contrôlée des informations sensorielles et contextuelles a permis d'évaluer leur complémentarité pour rendre compte de la performance des sujets et de l'évolution des stratégies de traitement de l'information visuelle en fonction de la répétition des stimulations.

7- BIBLIOGRAPHIE

- BIEDERMAN I. (1972). Perceiving real-world scenes. *Science*, Vol.177, pp.77-80.
- BIEDERMAN I., GLASS A.L. & STACY E.W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, Vol.97, pp.22-27.
- BIEDERMAN I. (1981). On the semantics of a glance at a scene. In *Perceptual organisation*, POMERANTZ J.R & KUBOVY M., pp.181-211. Ed : LEA, Hillsdale.
- BIEDERMAN I., MEZZANOTTE R.J. & RABINOWITZ J.C. (1982). Scene perception : Detection and judging objects undergoing relational violations. *Cognitive Psychology*, Vol.14, pp.143-177.
- BIEDERMAN I. (1987). Recognition-by-components : A theory of human image understanding. *Psychological Review*, Vol.94, pp.115-147.
- BIEDERMAN I. & JU G. (1988) Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, Vol.20, pp.38-64.
- BONDA E. (1991). L'identification des images naturelles : Rôle des processus attentionnels. Mémoire de DEA de Sciences Cognitives.
- BONNET C. (1986). Visual perception in context. *Cahier de Psychologie Cognitive*, Vol.6, pp.137-156.
- BONNET C. (1986). Manuel pratique de psychophysique, pp.15-19. Ed : Armand Colin, Paris.
- BONNET C. (1988). La perception visuelle des formes et du mouvement. *Intellectica*, Vol.5, pp.57-87.
- BONNET C. (1988). Les temps de traitement dans la perception visuelle des formes. Dans *Psychologie Cognitive, modèles et méthodes*, J.-P. CAVERNI, pp.309-328. Ed : P.U.G., Grenoble.
- BONNET C., GHIGLIONE R. & RICHARD J.-F. (1989). *Traité de psychologie cognitive I*. Ed : Dunod, Paris.
- BONNET C. (1995). Processus cognitifs dans la perception : La connaissance perceptive. *Rev. Neurol.*, Vol.151, pp.442-450.
- BRAJE W.L., TJAN B.S & LEGGE G.E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research*, Vol.35, pp.2955-2966.
- BRUCE V. & GREEN P.R. (1990). *Visual perception : Physiology, Psychology and Ecology*. Ed : LEA, Hove.

BURNOD Y. & OTTO I. (1994). Aquisition de l'information par l'observateur, modélisation des capacités adaptatives du système visuel humain. Science et défense 94, pp.355-364. Ed : Dunod, Paris.

CHEAL M. & LYON D.R. (1992). Attention in visual search : Multiple search classes. Perception & Psychophysics, Vol.52, pp.113-138.

DELORME A. (1982). Psychologie de la perception. Ed: Vigot, Québec.

DUBOIS D. & SPRENGER-CHAROLLES L.(1988). Perception/interprétation du langage écrit : Contexte et identification des mots au cours de la lecture. Intellectica, pp.113-146.

DUNKAN J. & HUMPREYS G.W. (1989). Visual search and stimulus similarity. Psychological Review, Vol.96, pp.433-456.

DUVAL-DESTIN M., MENU J.-P., PLANTIER J., BATEJAT D. & MEGARD P. (1989). Influence des fréquences spatiales sur l'extraction de cibles. Rapport CERMA 89-08.

D'YDEWALLE G. & VAN DAM G. (1988). Contextual information effects in text processing as a function of the type of test expected and received. Cahiers de Psychologie Cognitive, Vol.8, pp.139-152.

FIELD D.J., HAYES A. & HESS R.F. (1993). Contour integration by the human visual system : Evidence for a local "association field". Vision Research, Vol.33, pp.173-193.

FIORENTINI A., MAFFEI L. & SANDINI G. (1983). The role of higher spatial frequencies in face perception. Perception, Vol.12, pp.195-201.

FODOR J.A. (1986). La modularité de l'esprit : Essai sur la psychologie des facultés. Ed : Éditions de Minuit, Paris.

GEGENFURTNER K.R. (1996). The contribution of color to the recognition of natural scenes for normal and color deficient observers. Optical Society Meeting.

GINSBURG A.P. (1978). Visual information processing based on spatial filters constrained by biological data. Technical Report AMRL-TR-78-129.

GRAHAM N. (1981). Spatial frequency channels. In Perceptual organisation, POMERANTZ J.R & KUBOVY M., pp.23-25. Ed : LEA, Hillsdale.

HARMON L.D. (1973). The recognition of faces. Scientific American, Vol.227, pp.71-82.

HARMON L.D. & JULESZ B. (1973). Masking in visual recognition : Effect of two-dimensional filtered noise. Science, Vol.180, pp.1194-1197.

JOHANNES S., WIERINGA B.M., MATZKE M. & MÜNTE T.F. (1996). Hierarchical visual stimuli : electrophysiological evidence for separate left hemispheric global and local processing mechanisms in humans. Neuroscience letters, Vol.210, pp.111-114.

KAHNEMAN D. & HENICK A. (1981). Perceptual organisation and attention. In Perceptual organisation, POMERANTZ J.R & KUBOVY M., pp.181-211. Ed : LEA, Hillsdale.

KINCHLA R.A. & WOLFE J.M. (1979). The order of visual processing : "top-down", "bottom-up" or "middle-out". Perception & Psychophysics, Vol.25, pp.225-231.

KIRKPATRICK-STEGER K., WASSERMAN E.A & BIEDERMAN I. (1996). Effect of spatial rearrangement of object components on picture recognition in pigeons. Journal of the Experimental Analysis of Behavior, Vol.65, pp.465-475.

LAGASSE L. (1993). Effects of good form and spatial frequency on global precedence. Perception & Psychophysics, Vol.53, pp.89-105.

LOFTUS G.R. & MACKWORTH N.H. (1978). Cognitive determinants of fixation location during picture viewing. Journal of Exp. Psychol. : Human Perception and Performance, Vol.4, pp.565-572.

LOWE D.G. (1987). Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence, Vol.31, pp.355-395.

MALLAT S. (1989). Multifrequency channel decomposition of images and wavelet models. I.E.E.E. Transaction on acoustic speech and signal processing, Vol.37, pp.2091-2110.

MARR D. (1982). Vision : A computational investigation into the human representation and processing of visual information. Ed : Freeman, San Fransisco.

NAVON D. (1983). How many trees does it take to make a forest ? Perception, Vol.12, pp.239-254.

OLIVA A. & SCHYNS P.G. (1994). Analyse multi-échelle de la perception de scènes. Présenté aux 7^{èmes} journées des Neurosciences et Sciences de l'Ingénieur, Chamonix.

OLSHAUSEN B.A., ANDERSON C.H. & VAN ESSEN D.C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. The journal of Neurosciences, Vol.13, pp.4700-4719.

PALMER S. (1975). The effect of contextual scenes on the identification of objects. Memory & Cognition, Vol.3, pp.519-526.

PAQUET L. (1992). Global and local processing in nonattended objects : A failure to induce local processing dominance. Journal of Exp. Psychol. : Human Perception and Performance, Vol.18, pp.512-529.

PAQUET L. & MERIKLE P.M. (1984). Global precedence : The effect of duration. Canadian Journal of Psychology, Vol.38, pp.45-53.

PARKER D.M., LISHMAN J.R & HUGHES J. (1996). Role of coarse and fine spatial information in face and object processing. Journal of Exp. Psychol. : Human Perception and Performance, Vol.22, pp.1448-1466.

PLANTIER J. & MENU J.-P. (1993). Analyse des contrastes locaux de luminance dans les images complexes. Rapport CERMA 93-12.

ROSSI A.F. & PARADISO A.M. (1995). Feature-specific effects of visual search attention. *Vision Research*, Vol.35, pp.621-634.

SANOCKI T. & ODEN G.C. (1991). Perceptual adjustments on representations of familiar patterns: Change over time and relational features. *Perception & Psychophysics*, Vol.50, pp.28-44.

SANOCKI T. (1993). Time course of object identification : Evidence for global-to-local contingency. *Journal of Exp. Psychol. : Human Perception and Performance*, Vol.19, pp.878-898.

SCHWEICH M., SCREIBER A.-C., ROUSSET S., BRUYER R. & TIBERGHIE G. (1991). Effect of the meaning of visual context on semantic processing of famous faces. *Cahier de Psychologie Cognitive*, Vol.11, pp.55-71.

SEKULER R. & BLAKE R. (1990). Knowledge and perception. In *Perception*, pp.424-450. Ed : McGraw-hill international, Singapore.

SERGENT J. (1986). Microgenesis in face processing. In *Aspects of face processing*, ELLIS H.D., JEEVES M.A., NEWCOMBE F. & YOUNG A., pp.17-73. Ed : Martian Nighoff, Dordrecht.

STANOVICH K.E. (1980). Toward an interactif compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, Vol.16, pp.32-71.

TANAKA K. (1993). Neuronal mechanism of object recognition. *Science*, Vol.262, pp.685-688.

THEEUWES J. & KOOI F.L. (1994). Parallel search for a conjunction of contrast polarity and shape. *Vision Research*, Vol.34, pp.3013-3016.

THOMSON P. (1980). Margaret Thatcher : A new illusion. *Perception*, Vol.9, pp.483-484.

THORPE S. (1988). Traitement d'images chez l'homme. *Techniques et Sciences Informatiques*, Vol.7, pp.517-525.

THORPE S. (1995). La reconnaissance visuelle : de la rétine au cortex inférotemporal. *Revue de Neuropsychologie*, Vol.5, pp.389-410.

THORPE S., FIZE D. & MARLOT C. (1996). Speed of processing in the human visual system. *Nature*, Vol.381, pp.520-522.

TIBERGHIE G. (1986). Context and cognition. *Cahier de Psychologie Cognitive*, Vol.6, pp.105-119.

TIBERGHIE G. (1988). Perception et mémoire des visages. *Intellectica*, pp.89-111.

TREISMAN A. & GELADE G. (1980). A Feature Integration Theory of attention. *Cognitive Psychology*, Vol.12, pp.97-136.

TREISMAN A. (1986). Feature and objects in visual processing. *Scientific American*, Vol.255, pp.106-115.

TREISMAN A. & GORMICAN S. (1988). Feature analysis in early vision : Evidence from search asymmetries. *Psychological Review*, Vol.95, pp.15-48.

TSAL Y. & LAVIE N. (1993). Location dominance in attending to color and shape. *Journal of Exp. Psychol. : Human Perception and Performance*, Vol.19, pp.131-139.

UTTAL W.R., BARUCH T. & ALLEN L. (1995). The effect of combinations of image degradations in a discrimination task. *Perception & Psychophysics*, Vol.57, pp.668-681.

UTTAL W.R., BARUCH T. & ALLEN L. (1995). Combining image degradations in a recognition task. *Perception & Psychophysics*, Vol.57, pp.682-691.

WANG Q., CAVANAGH P. & GREEN M. (1994). Familiarity and pop-out in visual research. *Perception & Psychophysics*, Vol.56, pp.495-500.

WOLFE J.M., BENNETT S.C. (1997). Preattentive object files : Shapeless bundles of basic features. *Vision Research*, Vol.37, pp.25-43.

WOLFE J.M., FRIEDMAN-HILL S.R. & BILSKY A.B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception & Psychophysics*, Vol.55, pp.537-550.

WURM L.H., LEGGE G.E, ISENBERG L.M & LUEBKER A. (1993). Color improves recognition in normal and low-vision. *Journal of Exp. Psychol. : Human Perception and Performance*, Vol.19, pp.899-911.

YANTIS S. & JOHNSON D.N. (1990). Mechanisms of attentional priority. *Journal of Exp. Psychol. : Human Perception and Performance*, Vol.16, pp.812-825.