

DEA de sciences cognitives de Grenoble

Interactions entre les propriétés globales et locales des scènes naturelles

Yann-Aël LE BORGNE

Rapport de stage

Réalisé au LIS (Laboratoire des Images et Signaux) et
au LPNC (Laboratoire de Psychologie et de Neurocognition)

Encadrement : Nathalie GUYADER & Alan CHAUVIN

Juin 2003

Introduction	3
I) Catégorisation de scènes naturelles.....	4
1) Les scènes naturelles.....	4
2) Représentation fréquentielle	4
a) Séparation phase / amplitude	4
b) Spectre d'énergie	6
3) Exploitation du spectre d'énergie dans l'investigation du système visuel	7
a) Aspects neurophysiologiques.....	7
b) Modélisation	7
c) Validité psychoexpérimentale.....	7
4) Catégorie globale vs catégorie locale	8
II) Investigation des interactions global/local dans les scènes	10
1) Choix du protocole expérimental.....	10
Type des catégories	10
Paradigme d'amorçage.....	10
Taille des images globales et locales.....	10
Sélection des images	12
2) Expérience 1 : Influence du global sur le local.....	13
Hypothèses	13
Méthode.....	13
Résultats	15
3) Expérience 2 : Influence du local sur le global.....	17
Hypothèses	17
Méthode.....	17
Résultats	18
4) Discussion sur les résultats obtenus.....	20
a) Variabilité intergroupe	20
b) Interprétations des résultats.....	22
Conclusion.....	23
Annexe A : Fenêtrage de Butterworth	25
Annexe B : Masque dynamique.....	27
Annexe C : Images choisies	28
Annexe D : Ecrémage et distribution.....	33
Annexe E : Résultats et statistiques.....	37
Bibliographie	41

Introduction

Cette étude se place dans le cadre de la recherche sur la catégorisation d'images de scènes naturelles, adoptant à la fois une approche de type 'ingénieur', dont le but est la classification automatique d'images de scènes naturelles, et une approche de type 'psychologie expérimentale', dont le but est d'investiguer les processus mis en œuvre lors de la catégorisation de scènes naturelles par notre système visuel. Cette approche parallèle est permise par la collaboration de deux laboratoires au sein desquels s'est déroulé mon stage, le Laboratoire des Images et Signaux (LIS) et le Laboratoire de Psychologie et de Neurocognition (LPNC). L'intérêt de cette 'passerelle' interdisciplinaire est la complémentarité des recherches menées sur la catégorisation de scènes visuelles par les deux laboratoires, permettant à chacun d'orienter ses axes de recherches en fonction des connaissances acquises dans l'un des deux domaines.

Le groupe 'Signaux, Images et Communication' (SIC) du LIS œuvre dans la mise au point d'un classifieur automatique de scènes naturelles, dont les différents modules de traitement visent à simuler les caractéristiques connues de notre système visuel. Ce groupe de travail investigate également les améliorations possibles de ce classifieur par l'étude des propriétés statistiques des scènes naturelles.

Parallèlement, les travaux menés au LPNC dans le cadre de la recherche sur les scènes naturelles ont pour objectif de comprendre les mécanismes régissant le système visuel dans l'exploration et la reconnaissance de scènes.

Les récentes études menées sur les scènes naturelles ont montré que leurs spectres d'énergie possèdent des propriétés très intéressantes, qui permettent de caractériser le type d'environnement auquel appartient une scène. L'analyse des propriétés des spectres d'énergie fait partie des grands axes de recherche du groupe SIC, et les résultats probants obtenus par le classifieur, ainsi que la plausibilité biologique des traitements computationnels effectués, ont conduit à tester la portée de cette approche dans le cadre de la catégorisation de scènes chez l'homme. L'objectif de mon stage était de poursuivre l'investigation du rôle de l'information portée par le spectre d'énergie dans la catégorisation de scènes naturelles, en étudiant les interactions entre les propriétés globales et locales des spectres d'énergie dans les scènes naturelles.

La première partie de ce rapport est consacrée à la présentation du spectre d'énergie comme descripteur de catégories de scènes naturelles. Elle sera suivie de la partie expérimentale, consistant à étudier les influences réciproques des échelles globales et locales dans la catégorisation de scènes naturelles.

1) Catégorisation de scènes naturelles

1) Les scènes naturelles

Le terme de scène naturelle se réfère ici à une vue du monde réel, typique de l'environnement humain ; les catégories étudiées sont principalement des plages, des villes, des forêts, des montagnes, des routes et des intérieurs. Le choix de ce type de stimulus est lié à sa nature 'écologique', dans le sens où il correspond à notre environnement quotidien, et qu'il constitue un point de départ logique dans l'étude du système visuel humain, puisque c'est à partir de ce genre de stimulus que notre système visuel se structure. L'étude des scènes naturelles permet donc l'analyse du système visuel dans des conditions perceptives réalistes, ce qui est l'objectif recherché ici.

Malgré la complexité et la variabilité possibles des structures des différents types d'environnements naturels, de récents résultats expérimentaux appuient fortement l'hypothèse que la catégorisation de scènes ne requiert pas nécessairement l'analyse fine du stimulus visuel. La catégorisation d'une scène est en effet possible à partir de la seule information basse fréquence d'un stimulus, dans lequel l'identification des objets le composant n'est pas possible [OLI94]. De plus, de nombreux résultats expérimentaux ([OLI94], ou [HUG96] pour une revue) suggèrent que notre système visuel opérerait l'analyse d'un stimulus selon une stratégie par défaut de type 'coarse-to-fine', c'est-à-dire dans laquelle l'information serait traitée selon un ordonnancement basses fréquences/hautes fréquences. Ces hypothèses se basent sur des résultats de la neurophysiologie qui montrent que l'information basse fréquence d'un stimulus visuel est transmise plus rapidement que l'information haute fréquence, et sur des données de psychologie expérimentale qui suggèrent l'existence de 'canaux' fréquentiels, dans lesquels l'information basse fréquence serait traitée prioritairement [HUG96].

D'après ce modèle 'coarse-to-fine', l'information de basse fréquence amorcerait ainsi une première reconnaissance, à partir de la structure grossière des éléments dans l'image, qui serait ensuite validée ou rejetée par l'analyse fine du stimulus. La recherche d'invariant de scènes à un niveau global, permettant de caractériser différents types d'environnements visuels, a conduit à porter un intérêt très particulier au spectre d'énergie des scènes naturelles. Ce dernier s'affranchit en effet des variabilités structurales très importantes à l'intérieur d'une même catégorie (Les arrangements spatiaux des immeubles dans une ville sont par exemple infiniment nombreux, quel prototype structural pourrait en rendre compte ?). Mais il est cependant suffisamment informatif pour distinguer divers environnements naturels. De plus, le spectre d'énergie sert de base à certains traitements corticaux, et l'information qu'il véhicule jouerait un rôle important dans le processus de catégorisation de scènes naturelles par le système visuel (voir section 3).

2) Représentation fréquentielle

a) Séparation phase / amplitude

Du point de vue du traitement du signal, un stimulus visuel statique équivaut à une distribution spatiale de luminance (intensité lumineuse par unité de surface) et de chrominance (longueur d'ondes). Nous

considérons dans cette étude seulement des images carrées en niveau de gris. Mathématiquement, un stimulus visuel achromatique est une fonction $i(x,y)$ indiquant la luminance au point de coordonnées (x,y) . La théorie du signal permet de décomposer le stimulus en une somme de fonctions sinusoïdales complexes, dont les coefficients complexes $I(f_x, f_y)$ sont obtenus par la transformée de Fourier discrète :

$$I(f_x, f_y) = \sum_{x,y=0}^{N-1} f(x,y) e^{-2j\pi(f_x x + f_y y)}$$

où $f(x,y)$ est l'image $i(x,y)$ a subi un fenêtrage de Butterworth (voir annexe A pour les raisons de ce fenêtrage). Les intervalles de définitions des f_x et f_y sont tous deux $[-0.5 ; 0.5]$ après normalisation des fréquences par rapport à la taille de l'image $i(x,y)$ d'origine, et N est le nombre de pixels par côté dans l'image d'origine.

La fonction complexe $I(f_x, f_y)$ peut être réécrite en coordonnées polaires par deux fonctions $A(f_x, f_y)$ et $\Phi(f_x, f_y)$ qui correspondent respectivement au spectre d'amplitude (SA) et au spectre de phase (SP) de la transformée de Fourier :

$$I(f_x, f_y) = A(f_x, f_y) e^{j\Phi(f_x, f_y)}$$

Le spectre d'énergie (SE) correspond au spectre d'amplitude élevé au carré, et caractérise donc le même type d'information à un facteur de puissance près. Dans notre cadre de recherche, spectre d'amplitude et spectre d'énergie renvoie au même type d'information qualitative, c'est à dire la caractérisation des orientations dans l'image. Ceci est illustré ci-dessous :

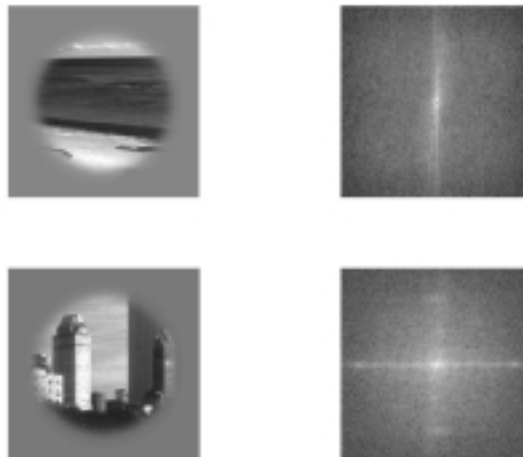


Figure 1 : Exemple de spectre d'énergie d'une image de plage et d'une image de ville

Sur ces exemples, le spectre d'énergie est représenté avec une échelle logarithmique pour une meilleure visualisation. Dans une plage, les orientations sont principalement horizontales. Cela se traduit dans le SE par de fort coefficients sur l'axe des ordonnées, qui correspondent à des sinusoïdes dont l'orientation est verticale. Pour l'image de ville, l'observation inverse se produit.

Le spectre de phase $\Phi(f_x, f_y)$ encode les contours précis dans l'image, comme l'illustre la figure suivante, qui montre à partir des deux images de plage et de ville, les images obtenues par les permutations des SA :

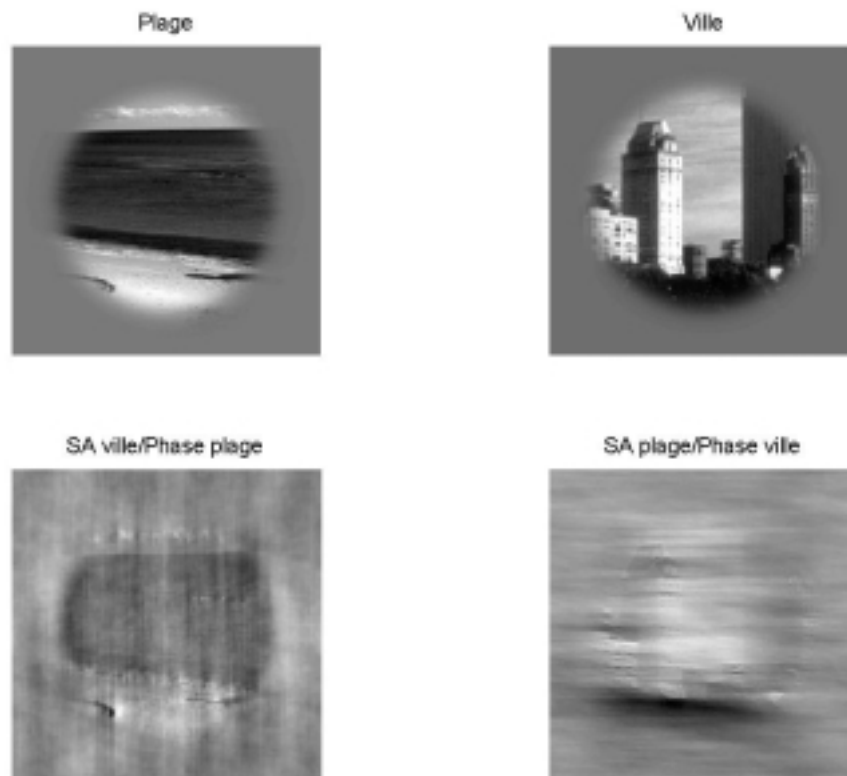


Figure 2 : permutation du SA de deux images : l'une de ville et l'autre de plage

La phase est très dépendante de la structure et des spécificités de l'image, ainsi que des positions absolues des composants de l'image. Elle contient une information plutôt relative aux propriétés locales de l'image.

b) Spectre d'énergie

Nous procédons ici à une revue rapide des propriétés du spectre d'énergie qui appuient son intérêt dans la catégorisation de scènes naturelles :

- La statistique sur les scènes naturelles montre que les spectres d'énergie de différentes catégories sont approximables par des fonction paramétrées en $1/f$, dont les paramètres sont caractéristiques de chaque catégorie [OLI01].
- Le spectre d'énergie d'une scène est la somme du spectre d'énergie moyen des composants de la scène, indépendants de leur position absolue, et d'un terme qui correspond à la position relative des composants entre eux [GUY02].
- Le choix de descripteurs appropriés issus du spectre d'énergie permet de simplifier le traitement d'opérations tels que les zooms, les rotations, ou l'inversion droite-gauche [GUY02].

3) Exploitation du spectre d'énergie

a) Aspects neurophysiologiques

L'image rétinienne d'un stimulus visuel est projetée sur le cortex visuel primaire, où elle est décomposée en un certains nombres de primitives. L'étude des cellules de V1 (Hubel & Wiesel) a montré l'existence de cellules (cellules simples et complexes de la sous-couche 4c) sélectives aux fréquences et aux orientations des stimuli visuels. Le rôle de ces cellules est modélisable par des ondelettes de Gabor [LEE96], qui échantillonnent le spectre d'énergie d'une image en log-polaire. Les données de la neurophysiologie mettent en évidence des sélectivités plus étroites de ces cellules au niveau des basses fréquences que des hautes fréquences.

b) Modélisation

Ceci se traduit au niveau de la modélisation par un échantillonnage en log-polaire du spectre d'énergie en fonction de certaines fréquences et certaines orientations. Le spectre d'énergie de l'image est échantillonné après des prétraitements de l'image qui consistent à égaliser les contrastes et à opérer un blanchiment spectral comme le fait la rétine (Hérault & Beaudot, cité par [GUY02]). Cet échantillonnage fournit une signature spectrale de l'image, sous forme de matrice, composée de l'énergie renvoyée en sortie des filtres pour l'ensemble des orientations et des fréquences choisies.

Appliquée à un ensemble d'images d'une même catégorie, cette procédure d'échantillonnage permet par moyennage d'obtenir une signature spectrale de référence de la catégorie. L'appartenance d'une image à une catégorie s'effectue ensuite par l'évaluation de la distance entre la signature spectrale de l'image et celle de référence des diverses catégories. La détermination du barycentre de chaque catégorie fait encore l'objet de recherches (moyennes des signatures des images d'une catégorie, convergence de cartes de Kohonen, recherche des plans principaux de la distribution des points, etc...), et le calcul de distance entre la signature de l'image et la signature de référence est également investigué (euclidienne, de Minkowski, de Mahalanobis, aux plans principaux) [GUY02].

c) Validité psychoexpérimentale

L'influence spécifique de l'information d'amplitude et de phase a été testée chez le sujet humain dans des tâches de catégorisation rapides de scènes naturelles ([CGMH02], et Guyader, rapport en cours). Dans ces expériences, les sujets devaient catégoriser des images de plage et de ville le plus rapidement possible. Un protocole d'amorçage (identique à celui que nous utilisons dans la seconde partie de ce rapport) permettait d'étudier l'influence de la perception implicite d'une image (l'amorce) sur la catégorisation explicite d'une image de plage ou de ville (la cible). La construction d'images 'hybrides', dont le spectre d'amplitude était issu d'une image de plage (respectivement de ville) et dont le spectre de phase était issu d'une image de ville (respectivement de plage), permettait de tester l'influence implicite de l'information de phase et d'amplitude sur la catégorisation des images cibles. Il a été montré que les amorces hybrides dont le SA était congruent à celui de la cible (c'est à dire que

le SA de l'amorce était issu d'une image différente mais de même catégorie que celle la cible) facilitaient le traitement de la cible (les temps de catégorisation étaient plus courts) par rapport à la condition où les amorces n'étaient pas hybrides et appartenaient à l'autre catégorie que celle de la cible (condition non congruente). En revanche, les amorces hybrides dont le SP était congruent à la cible (C'est à dire dont le SP était issu d'une image différente mais de même catégorie que celle de la cible) ne facilitaient pas le traitement de la cible par rapport à la condition non congruente. Ceci démontre l'importance de l'information portée par le spectre d'amplitude d'une image de scène naturelle dans la catégorisation rapide de celle-ci. C'est également un résultat contre intuitif dans la mesure où les images hybrides (cf. figure 2) ne sont catégorisables explicitement qu'à partir de leur information de phase. Le spectre d'amplitude a donc un rôle implicite important dans le processus de catégorisation de scènes naturelles, et offre de nouvelles voies de recherches dans la compréhension des traitements effectués par notre système visuel dans le traitement de l'information rétinienne. Pour l'instant, nous avons considéré la catégorisation de scènes naturelles à l'échelle globale, avec le spectre d'amplitude comme descripteur. Cependant, la catégorie d'une image n'est souvent pas exclusive, du fait de la variété d'environnements qui peuvent composer une vue. C'est ce genre de situation sur lequel nous allons maintenant nous pencher.

4) Catégorie globale vs catégorie locale

Considérons l'exemple ci-contre. Une vue du mont St Michel rentre-t-elle plutôt dans la catégorie ville ou la catégorie plage ? Les deux, évidemment, et ce genre de situation est plus que fréquent. D'une manière générale, un environnement a une composition variée, et les catégories que nous avons mises en évidence précédemment n'ont qu'une réalité locale, ou globale seulement si toutes les catégories locales s'accordent de façon presque unanime. Ainsi, l'analyse d'un environnement doit être menée à l'échelle locale pour prendre en compte la diversité de la composition de la vue globale. Dans notre modèle, selon lequel la catégorie d'une scène serait principalement inférée par la signature spectrale du signal visuel, cette stratégie revient à décomposer l'image globale en imagettes de tailles réduites, à calculer la signature spectrale de chaque imagette, puis à évaluer localement la catégorie de chaque imagette. La catégorie globale, si catégorie globale il y a, serait celle à laquelle une majorité d'imagettes appartiennent.



Figure 3 : Une scène visuelle 'composite', localement de type ville et de type plage

Du point de vue du traitement automatique, comme l'on peut s'y attendre, cette approche accroît le degré de précision des classements. Mais, dans notre cadre de recherche, nous sommes également intéressés par la validité de cette approche pour le système visuel humain. Or cette double hypothèse d'une représentation de la catégorie d'une scène par sa signature spectrale globale, ou par

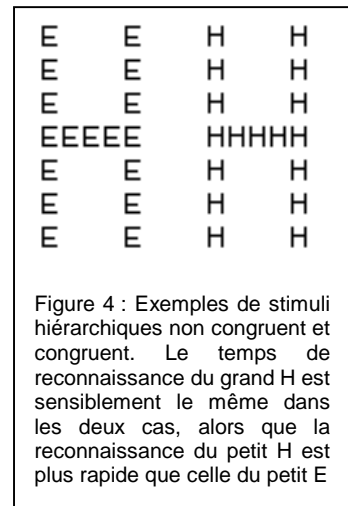
l'évaluation des différentes signatures spectrales de parties locales de la scène, n'est pas sans rappeler les stimuli hiérarchiques.

Initiées par Navon en 1977, les expériences sur les stimuli hiérarchiques investiguent les interactions entre perception globale et locale. Les stimuli hiérarchiques sont des stimuli dans lesquels l'agencement de formes locales fait émerger une forme globale (figure 4). Deux types de conditions sont étudiés, un cas où la grande forme et les petites formes sont identiques (congruence), et un cas où la grande forme et les petites formes sont différentes (non congruence). On teste le temps de reconnaissance de la grande forme ou des petites formes dans ces deux conditions. Deux phénomènes sont observés :

- La précedence globale : Le temps de reconnaissance de la forme globale est plus rapide que celui de la forme locale.
- L'interférence globale : L'information globale interfère sur le traitement de l'information globale, et non l'inverse (le temps de reconnaissance de la petite forme est plus long dans la condition non congruente que dans la condition congruente, alors que le temps de reconnaissance de la grande forme est sensiblement le même dans les deux conditions).

Ces rapports entre perception globale et perception locale n'ont jamais été étudiés sur des scènes naturelles, qui sont des stimuli beaucoup plus complexes. On peut se demander si ce que nous observons dans les expériences de stimuli hiérarchiques peut se généraliser au cas des scènes naturelles, avec le SA comme alphabet.

C'est l'objectif de cette deuxième partie, dans laquelle nous étudions dans les scènes naturelles l'effet de la perception globale sur la perception locale (expérience 1), et l'effet de la perception locale sur la perception globale (expérience 2).



II) Investigation des interactions global/local dans les scènes

1) Choix du protocole expérimental

Type des catégories

Nous avons choisi pour ces expériences les catégories 'plages' et 'villes', du fait de leurs signatures spectrales très distinctes qui favorisera les conditions congruentes et non congruentes entre global et local.

Paradigme d'amorçage

L'interaction global/local est étudiée par le biais d'un paradigme d'amorçage. Le principe consiste à étudier l'influence d'un stimulus visuel traité implicitement (l'amorce) sur le temps de catégorisation explicite d'une scène (la cible). Pour tester l'influence du global sur le local, on utilise comme amorce une image 'globale', relativement à l'image cible qui sera une partie d'image 'globale'. Nous reviendrons sur la relativité global/local dans la section suivante. Le test de l'influence de la perception locale sur la perception globale suit le même principe, en inversant la taille de l'image entre l'amorce et de la cible. La congruence entre amorce et cible est définie comme l'appartenance de l'imagette à la même catégorie que celle de l'image globale, avec des spectres d'amplitude pour l'image globale et pour l'imagette typiques de leurs catégories. La non congruence est définie comme l'appartenance de l'imagette à une image de l'autre catégorie que celle de l'image globale, et SA d'imagette et d'image globale sont typiques de leur catégories respectives.

Le traitement implicite de l'amorce est obtenu par l'utilisation d'un masque dynamique [BAC02], qui sépare la présentation de l'amorce de celle de la cible, et dont l'apparition très rapprochée dans le temps de la présentation de l'amorce permet de stopper l'analyse de l'image d'amorce avant qu'elle ne soit perçue de façon explicite. Le masquage dynamique consiste en la présentation successive d'une suite d'images contenant du bruit à différentes échelles spatiales, dont le but est de brouiller sur la rétine toutes les informations relatives à l'image d'amorce. Les images de bruits sont présentées successivement depuis l'échelle haute fréquence jusqu'à l'échelle basse fréquence. Les images correspondant à ce masque se trouvent en annexe B.

Taille des images globales et locales

La taille des images disponibles dans la base d'images du laboratoire est de 256*256 pixels. Affichée sur un écran 21 pouces avec une résolution de 1024*768, l'image a une taille angulaire de 9.7° par 9.7° .

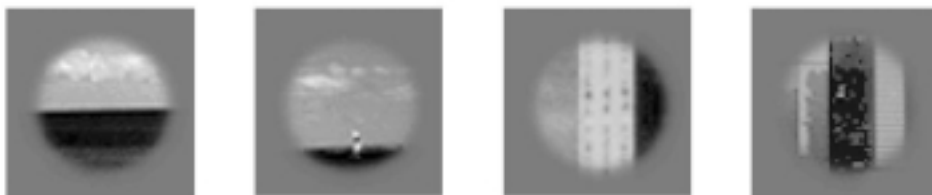


Taille d'une image de type 'global' telle qu'elle est affichée sur l'écran. L'image est vue à une distance de 80 cm.

La taille des imagettes (c'est ainsi que nous référerons à l'extrait local d'une image) a été définie comme la taille minimale permettant de vérifier les deux contraintes suivantes :

- Il fallait qu'elle permette une reconnaissance de l'image dont elle était extraite après le fenêtrage de Butterworth, qui réduit légèrement la dimension de l'imagette.
- Le spectre d'amplitude (SA) de l'imagette devait être similaire au SA prototypique de la catégorie dont elle était extraite.

Une taille de 64*64 pixels, correspondant à 1/16 d'une image globale, a d'abord été choisie.

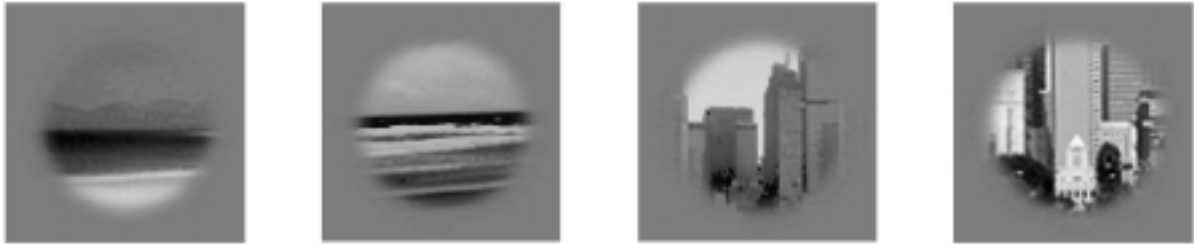


Exemples d'imagettes de plages et de villes au format 64*64 pixels

Mais les imagettes obtenues présentaient un biais qui nous a conduit à augmenter leur taille. En effet, la contrainte du choix d'imagettes dont le SA devait être similaire à celui du SA de sa catégorie imposait de ne choisir que des imagettes présentant des orientations horizontales (pour les plages) ou verticales (pour les villes). Dans une première série d'expériences menée avec des imagettes de taille 64*64, les sujets, dans un questionnaire post-expérimental, nous disaient avoir effectué leur choix non pas en fonction de la catégorie de l'imagette présentée, mais en fonction de l'orientation. Il y

avait donc un problème évident entre tâche prescrite et tâche effective qu'il nous était difficile de contrôler, et nous avons préféré augmenter la taille des imagerie afin d'éviter ce biais.

Une taille de 96*96 pixels a donc été choisie, soit 1/9 de l'image globale. Dans ces imagerie, la nature des contrastes reste globalement horizontale ou verticale, mais de façon beaucoup moins prononcée que précédemment, et les sujets, bien que relatant toujours cette caractéristique lors du questionnaire post-expérimental (mais ceci est normal puisqu'elle est inhérente aux catégories plage et ville), disaient avoir déterminé leur choix principalement par la reconnaissance de la catégorie à laquelle leur semblait appartenir l'imagerie.



Exemples d'imagerie de plages et de villes, au format 96*96 pixels

Un soin particulier a été donné à la sélection des imagerie pour s'assurer que les sujets détermineraient leur choix à partir de la catégorie de l'imagerie.

Sélection des images

Les images sont issues de la base d'image du laboratoire, qui contient environ 60 images dans chacune des catégories ville et de plage. Les images sont des vues d'ensemble, au format 256*256 pixels, avec 256 niveaux de gris.

Sélection des images globales :

Une fois les images de la base normalisées et fenêtrées, nous avons choisi pour neuf images pour chaque catégorie plage et ville. Le critère de sélection principal fut, après l'élimination des images qui avaient des variances trop faibles, l'aspect du spectre d'amplitude des images.

Ainsi, pour les images de plages, ont été retenues les neuf images qui avaient un SA présentant au mieux de fortes amplitudes selon l'orientation verticale. Ce type de caractéristique spectrale est ce qui correspond d'une manière générale aux scènes de plages. Pour les images de villes, le SA devait avoir pour caractéristique de présenter essentiellement de fortes amplitudes selon les orientations horizontales. D'une manière générale, le spectre d'amplitude des villes présente aussi des orientations verticales du fait des fenêtres ou des toits (contrastes selon des lignes horizontales dans l'image), mais les coefficients selon ces orientations sont moins marqués que ceux selon l'orientation horizontale. Ce renforcement des traits distinctifs entre SA des catégories ville et plage permet d'orthogonaliser les spectres, et d'éviter les interférences dans les conditions congruentes et non

congruentes que nous testerons. Les images choisies ainsi que leurs statistiques se trouvent en annexe C.

Sélection des images locales :

Il s'agit ensuite ici de choisir également neuf images dans chaque catégorie, au format 96*96 pixels. Une première sélection d'une vingtaine d'images dans chaque catégorie, ayant les SA voulus (même principe que pour les images globales) après fenêtrage, ainsi que des moyennes de luminosité autour de 128 et des variances similaires, a été effectuée.

Les images étant de format réduit et en intensité de gris, elles sont moins facilement catégorisables que les images globales. Afin de s'assurer qu'elles étaient facilement identifiables, nous les avons présentées à quatre sujets, en insérant dans le paquet des quarante images ainsi retenues quarante images 'distractrices', provenant de scènes d'intérieurs (qui peuvent être confondues lors d'une présentation rapide avec les images de villes et qui présentent un SA typique de ville) et de paysages champêtres (pouvant se confondre avec les images de plage et présentant un SA typique de plage). Chaque sujet devait classer en quatre catégories ville-plage-intérieur-champ le paquet d'images ainsi constitué, ceci le plus rapidement possible. A l'issue de ces présentations, les images de plages et de villes incorrectement classées ont été disqualifiées.

Sur les images restantes, nous avons repris les critères de l'aspect du SA, de la moyenne de luminosité et de la variance pour sélectionner les neuf images de chaque catégorie. Les images choisies ainsi que leur spectre se trouvent également dans l'annexe C.

2) Expérience 1 : Influence du global sur le local

Hypothèses

Nous attendons ici un effet de la congruence sur les temps de réaction (TR), et sur les pourcentages d'erreurs (PE) :

- $TR_{\text{congruent}} < TR_{\text{neutre}} < TR_{\text{noncongruent}}$
- $PE_{\text{congruent}} < PE_{\text{neutre}} < PE_{\text{noncongruent}}$

Les expériences de catégorisation de plages et de villes conduites précédemment par notre laboratoire ont obtenu des taux d'erreurs quasiment nuls, donc l'hypothèse de l'influence de la congruence sur le taux d'erreurs est moins forte que celle sur les temps de réactions.

Méthode

Participants : 34 étudiants en psychologie de l'université Pierre Mendès France de Grenoble. Tous les sujets avaient une vue normale ou corrigée, et ne connaissaient pas les hypothèses de l'expérience.

Stimuli : 54 paires d'images, chaque paire étant constituée d'une image d'amorce et d'une image test. Les images d'amorce sont des images globales, de 256*256 pixels, fenêtrées par un fenêtrage de

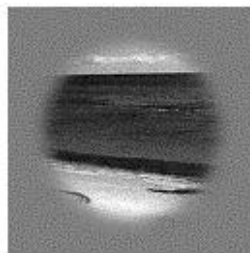
Butterworth. Elles sont en 256 niveaux de gris, avec des moyennes de luminosité autour de 128, et des variances similaires. Il y a 27 images d'amorce, 9 de plages, 9 de villes et 9 de bruits blancs (neutre). Les images tests sont des imagerie, au format 96*96 pixels. Il y a 18 imagerie en tout, divisées en deux catégories : 9 imagerie de plages et 9 imagerie de villes. Images et imagerie sont issues de la base d'images du laboratoire, qui contient environ une quarantaine d'images par catégories.

A chaque type d'image d'amorce (ville, plage, ou neutre) est associé une imagerie de la catégorie ville et une imagerie de la catégorie plage. Comme il y a 27 images d'amorce, on obtient ainsi 54 paires. L'imagerie associée à une image n'est pas extraite de cette image, ceci pour éviter une facilitation non contrôlée de la catégorisation de l'imagerie de part son appartenance à l'image d'amorce.

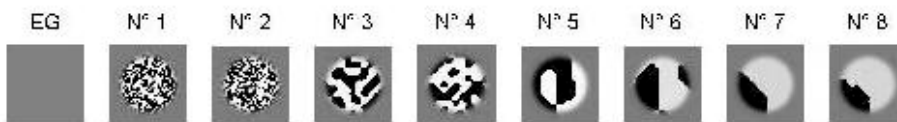
Pour la phase d'apprentissage, dix imagerie sont utilisées, cinq de villes et cinq de plages. Ces imagerie ne font pas parties des imagerie présentées lors de l'expérience.

Appareillage : Les stimuli sont affichés sur un écran 21 pouces à 100Hz. Les réponses sont obtenues grâce à un bouton poussoir situé sur un boîtier dédié connecté à une carte d'entrée/sortie de l'ordinateur. Les temps de réaction sont calculés à la milliseconde près.

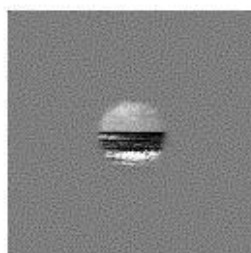
Procédure :



Présentation de l'amorce, 10ms



EG : Ecran gris, 30ms puis masque dynamique, 8 images, 20ms par image soit 160ms en tout



Présentation de la cible, 40ms

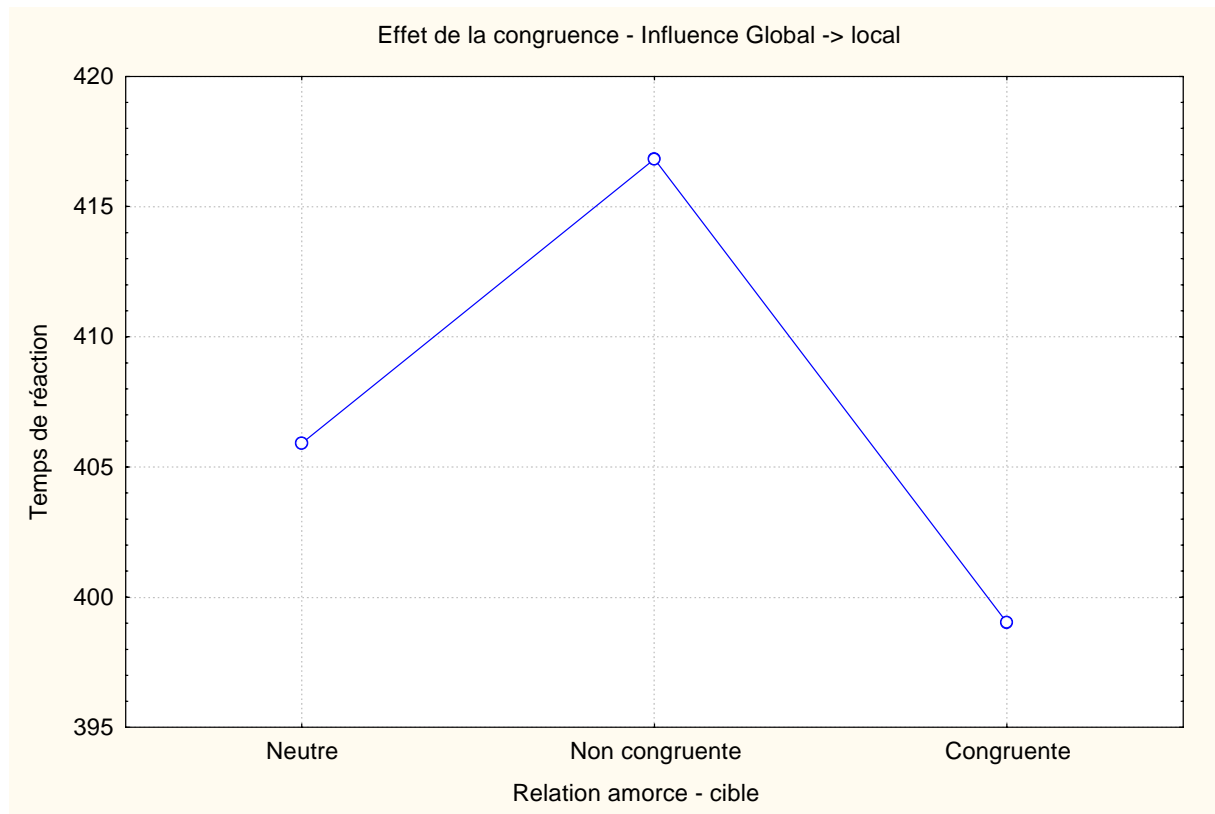
Exemple d'essai congruent de la catégorie plage

Chaque sujet est testé individuellement. Les sujets sont divisés en deux groupes. La tâche est de type go/no go. Le groupe P (14 sujets) a pour consigne d'appuyer sur le bouton poussoir quand il reconnaît une imagerie de plage, et de ne rien faire pour une imagerie de ville. Le groupe V (20 sujets) doit appuyer quand il reconnaît une imagerie de ville, et ne rien faire pour une imagerie de plage. Les sujets sont installés à environ 80cm de l'écran. L'expérience commence par une phase d'habituation à la tâche. Un écran indiquant les consignes est d'abord présenté au sujet, qui décide quand commencer après l'avoir lu. Les dix images de la phase d'apprentissage sont présentées dans un ordre aléatoire, pour un total de 18 essais. Elles sont précédées du masque, mais dans ces présentations il n'y a pas d'image d'amorce. Cette phase d'apprentissage a seulement pour but d'habituer le sujet au protocole expérimental. Un taux de réussite de 80% est nécessaire pour passer à la phase expérimentale. Ensuite, la phase expérimentale débute. Cette phase comprend deux sessions comprenant chacune la présentation des 54 essais, et diffèrent seulement dans l'ordre de présentations des essais qui est aléatoire. Pour chaque session, un nouvel écran rappelle les consignes, et le sujet choisi quand commencer. Un essai consiste en la présentation d'un point de fixation, suivie d'une présentation brève de l'image d'amorce (10ms), suivie d'un écran gris (30ms), suivie du masque (chaque image du masque est présentée 20ms, soit un temps de masquage total de 160ms ; La suite d'image est toujours la même), et enfin de l'image test (40ms). La durée totale de l'expérience est d'environ 15 minutes.

Résultats

Nous avons réalisé une ANOVA sur les temps de réaction et sur les pourcentages d'erreur. Nous réalisons une analyse canonique sur les trois facteurs que sont le groupe (variable inter, 2 modalités ville/plage), la session (variable intra, 2 modalités 1/2), et la congruence amorce-cible (variable intra, 3 modalités neutre/congruent/non congruent). Le pourcentage d'erreur (Fausses alarmes et omissions confondues) est de 1,13%, et nous ne l'étudions qu'à titre indicatif pour l'effet de la congruence du fait de sa faible importance. L'ensemble de ces résultats statistiques est présenté en annexe 3.

Effet de la congruence :



Temps de réaction moyens en fonction des conditions de congruence amorce-cible

Pour l'analyse des temps de réaction, nous obtenons un effet significatif de la congruence entre l'amorce et la cible, $F(2,64)=7.51$; $p<.001$. Un test post hoc de Newman-Keuls indique que le temps de catégorisation dans la condition congruente est plus court (399ms) que pour la condition non congruente (417ms), avec un $p<.0008$. Cette observation est appuyé par l'analyse des taux d'erreurs qui indique également un effet significatif de l'amorce ($F(2,64)=4.65$, $p<.013$), et le pourcentage d'erreurs est significativement plus faible pour la condition congruente (0.86%) que pour la condition non congruente (1.6%) (test de Newman-Keuls, $p<.016$).

Le temps de catégorisation pour la condition neutre (406ms) se place bien entre les temps de catégorisation pour les conditions congruentes et non congruentes. Mais la procédure de Newman-Keuls montre qu'il n'est significativement différent que de la condition non congruente ($p<.02$). La différence entre condition neutre et condition congruente n'est pas significative avec un $p<.14$.

De même, le taux d'erreur pour la condition neutre (0.96%) se place entre le taux d'erreurs de la condition congruente (0.86%) et la condition non congruente (1.6%), mais n'est significativement différent que de la condition non congruente (Newman-Keuls, $p<.016$).

Autres effets :

L'analyse des temps de réaction montre que la moyenne des temps de réaction en première session ($t = 417$ ms) est supérieure à celle de la seconde session ($t = 389$ ms), $F(1,32)=9.31$, $p<.004$, qui peut s'expliquer par une habitude à la tâche des sujets en deuxième session. Cet effet n'entre pas en

interaction avec l'amorce, donc il n'est pas gênant. Une phase d'apprentissage plus longue pourrait probablement le faire disparaître.

Bilan :

Ces résultats sont conformes à nos hypothèses. On observe une facilitation du traitement d'une image locale dont la catégorie d'appartenance et le SA sont congruents avec ceux de l'image globale, par rapport à une condition non congruente où l'image globale a une catégorie d'appartenance et un SA différent de ceux de l'image locale. Cette facilitation est observée à la fois dans les temps de réaction et dans le taux d'erreurs. On a ainsi mis en évidence un effet d'interférence globale dans le traitement local d'un stimulus de scène naturelle.

3) Expérience 2 : Influence du local sur le global

Hypothèses

Cf expérience 1.

Méthode

Participants : 33 étudiants en psychologie de l'UPMF de Grenoble. Tous les sujets avaient une vue normale ou corrigée, et ne connaissaient pas les hypothèses de l'expérience.

Stimuli : Les mêmes images que pour l'expérience 1 sont utilisées, sauf pour les bruits blancs : comme ici les images d'amorce sont des imasettes et les images cibles des images 'globales', les neuf images de bruits blancs sont remplacées par 9 imasettes de bruits blancs. On construit les 54 paires d'images suivant la même méthode que pour l'expérience 1.

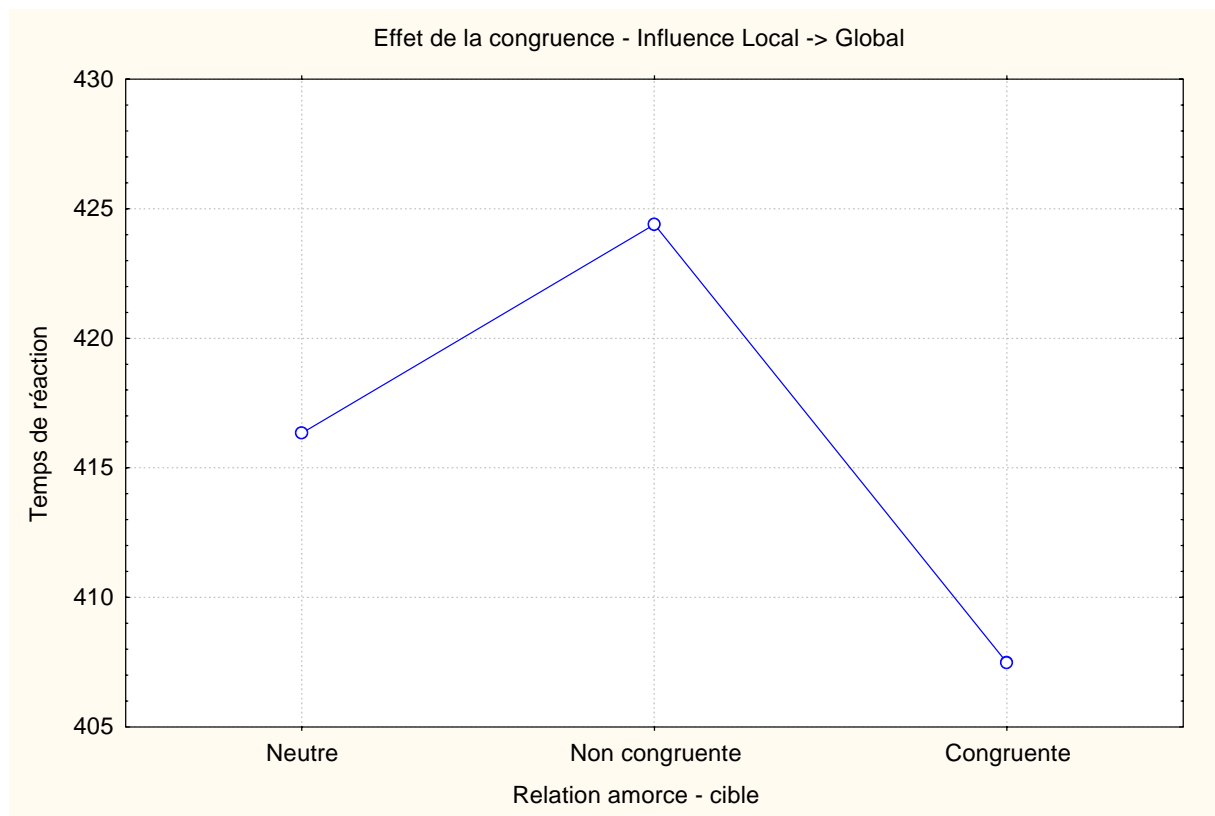
Appareillage : cf expérience 1.

Procédure : Le groupe P comporte ici 17 sujets et le groupe V 16 sujets. Les amorces sont maintenant des imasettes, et les cibles des images 'globales'. Les temps de présentation des amorces et des cibles ont été modifiés de manière à prendre en compte les différences de facilités perceptives entre les imasettes et des images globales. Le temps de présentation de l'image cible (globale) a été ramené à 20ms (c'est le temps de présentation utilisé pour des images cibles de ce format dans les autres études du laboratoire ayant le même protocole). Le temps de présentation de l'amorce a été augmenté car le stimulus d'amorce ayant dans cette expérience un format 9 fois inférieur à celui de la première expérience, son effet d'amorçage en est diminué. Il a été fixé à 30ms suite à une expérience pilote qui a montré que pour des temps de présentation inférieurs (20ms), l'amorce n'avait aucune influence sur la catégorisation.

Résultats

Le plan factoriel est ici le même que dans l'expérience 1. Nous avons également réalisé une ANOVA sur les temps de réaction et le pourcentage d'erreur. Le pourcentage d'erreur est ici un peu plus élevé, à 1,7% des réponses totales (Fausses alarmes et omissions confondues), et nous ne l'étudierons également qu'à titre indicatif. L'ensemble de ces résultats statistiques est présenté en annexe 3.

Effet de la congruence :



Temps de réaction moyens en fonction des conditions de congruence amorce – cible

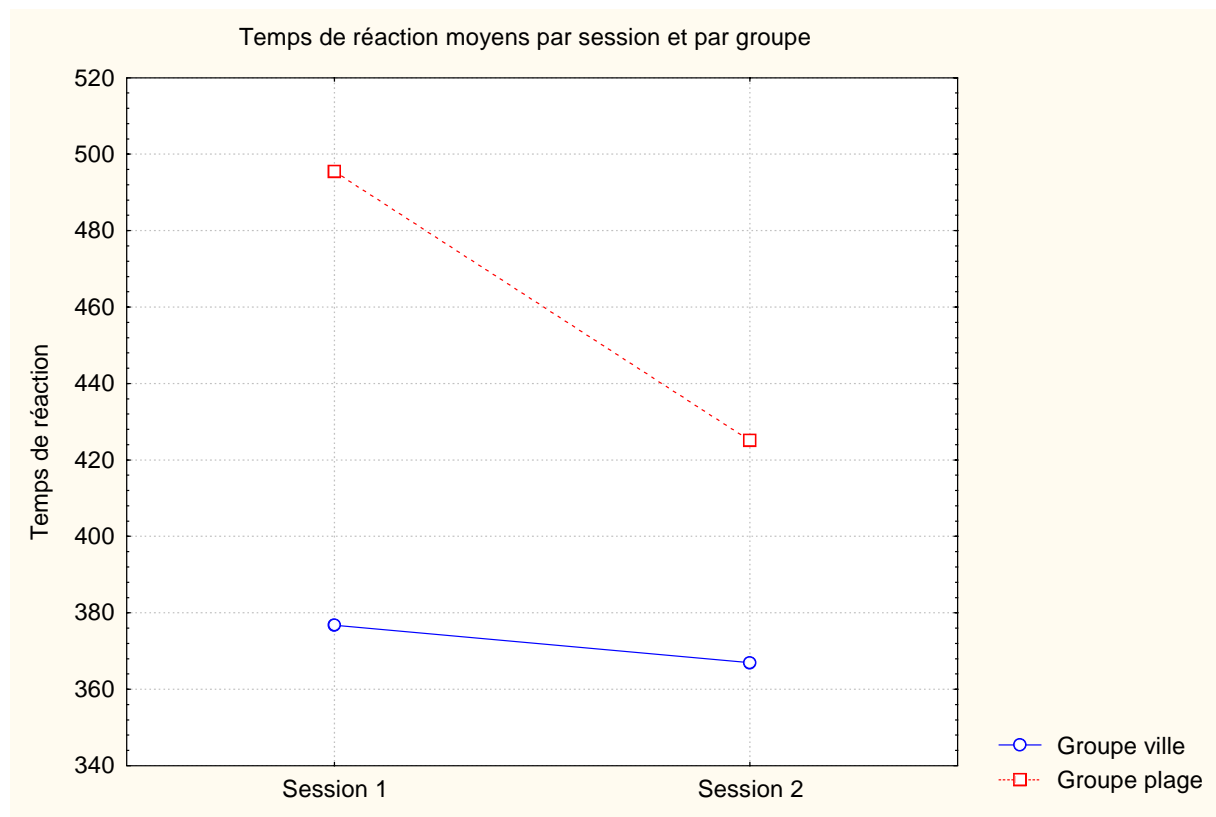
L'analyse des temps de réaction renvoie ici aussi un effet significatif de la congruence entre l'amorce et la cible, $F(2,62)=5.02$; $p<.009$. Les temps de catégorisation moyen des images locales sont plus rapides dans la condition congruente (40ms) que dans la condition non congruente (424ms) (test post-hoc de Newman-Keuls, $p<.0067$). Le temps de catégorisation moyen de la condition neutre (407ms) se place entre les temps de catégorisation moyens des conditions congruentes et non congruentes, mais les écarts entre condition neutre et les deux autres conditions ne sont pas significatifs.

Pour les résultats de l'analyse des pourcentage d'erreur, l'effet principal de la congruence n'est pas significatif, $F(2,62)=2.97$, $p<.058$.

Autres effets :

Comme dans l'expérience 1, le temps de catégorisation moyen lors de la première session est plus long (436ms) que lors de la deuxième session (396ms) ($F(1,31)=10.27$, $p<.003$). L'analyse du pourcentage d'erreurs nous montre également que le taux d'erreurs est plus important en première session (2,2%) qu'en deuxième session (1,1%) ($F(1,31)=10.4$, $p<.003$). Ceci est aussi probablement imputable à l'habituatation à la tâche. Il n'y a pas d'interaction entre la session et le type de congruence, donc cet effet principal de la session n'a pas d'implication sur l'effet de la congruence.

Enfin, nous observons que le temps de catégorisation des villes est beaucoup plus rapide (371ms) pour le groupe ville que pour le groupe plage (460ms) ($F(1,31)=13.9$, $p<.0008$). Cet effet est doublé d'une interaction entre le groupe et la session ($F(1,31)=5.86$, $p<.02$). Le tracé des temps de réaction par session et par groupe est illustré ci-après :



Variabilité entre les groupes

La variabilité des temps de réaction entre la session 1 et la session 2 est beaucoup plus marquée pour le groupe plage, ce qui explique l'interaction groupe-session. Nous analyserons plus précisément ces résultats d'effets du groupe dans la section suivante.

Bilan :

Les résultats obtenus sont ici aussi conformes aux hypothèses pour les temps de réactions : la perception globale d'une scène naturelle est plus rapide dans le cas où localement l'information est congruente avec l'information globale (mêmes catégories d'appartenance et SA typiques de cette

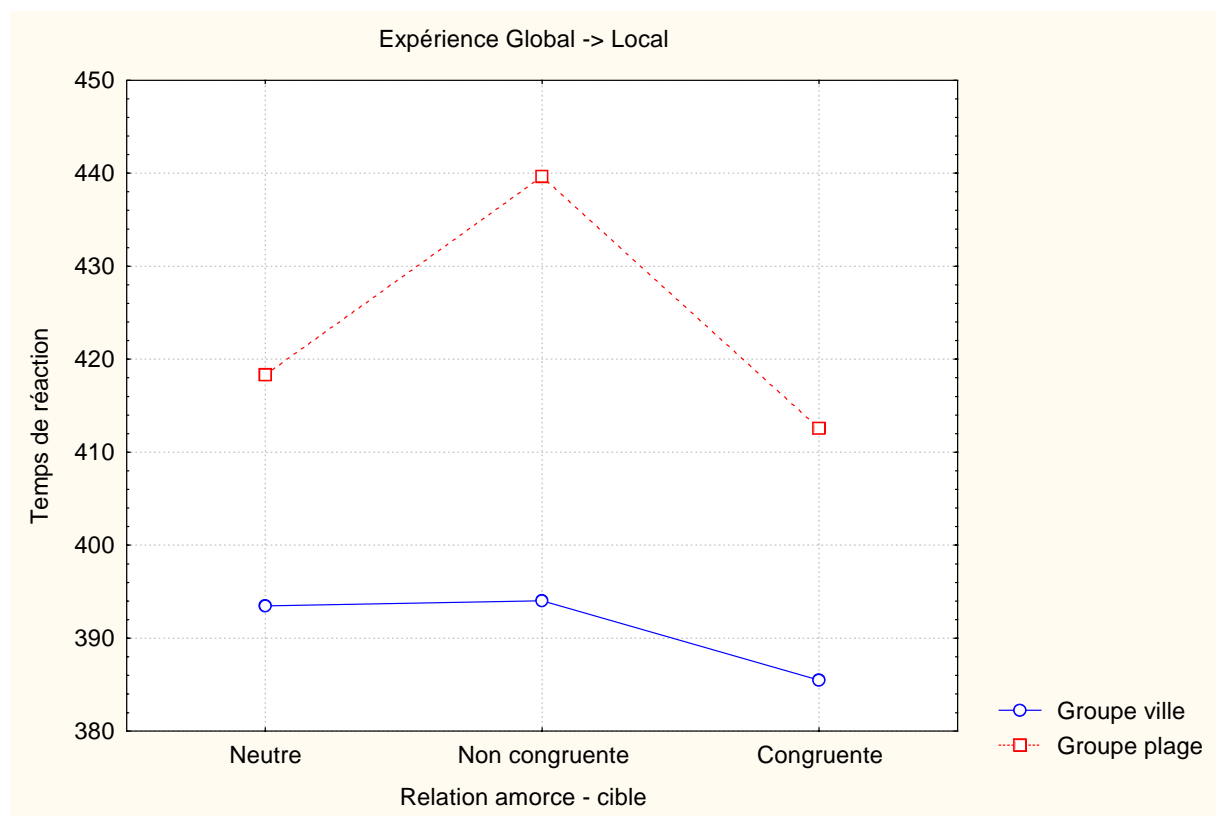
catégorie), que dans le cas où cette information est non congruente. Nous n'avons cependant pas obtenu d'effet de la congruence à travers les taux d'erreurs, mais cette hypothèse était moins forte du fait de leur faible taux. L'effet que nous avons obtenu ici est un effet d'interférence locale (influence de la perception locale sur la perception globale). C'est un résultat qui ne s'accorde pas avec ceux obtenus dans les expériences classiques sur les stimuli hiérarchiques, et nous reviendrons sur les conclusions possibles de cette singularité dans la section suivante.

4) Discussion sur les résultats obtenus

a) Variabilité intergroupe

Lors de l'analyse des résultats, nous avons confondu pour chacune des expériences les résultats du groupe 'ville' et du groupe 'plage'. Or il semble que l'effet de la congruence des SA entre l'amorce et la cible ne soit pas du même ordre entre ces deux groupes. L'étude de l'effet simple de l'amorce pour chaque groupe montre en effet des différences qui méritent attention.

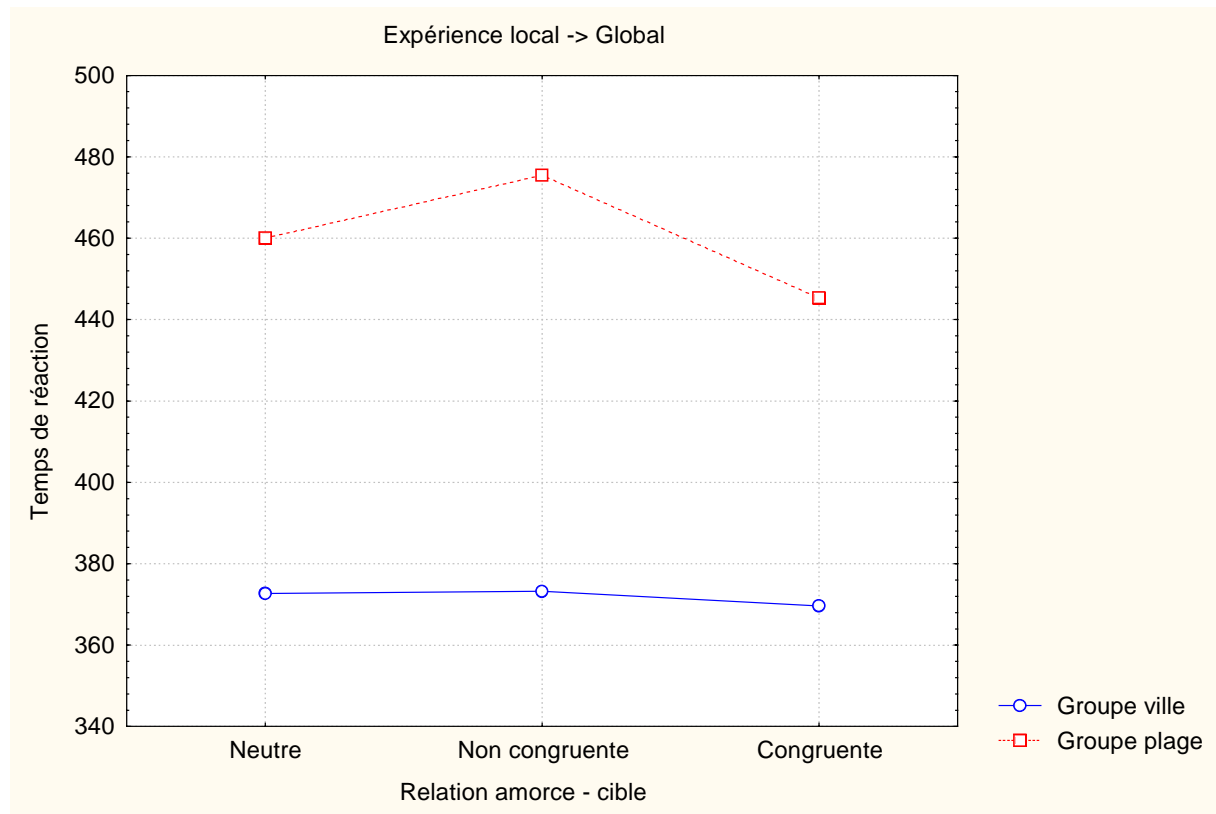
Le tracé des temps de réaction par groupe en fonction du type de congruence pour l'expérience 1 est illustré ci-après :



Une sensibilité à la congruence différente entre les groupes ville et plage

Un contraste sur le facteur de congruence pour le groupe ville nous montre qu'il n'y a pas d'effet de facilitation de la condition de congruence pour ce groupe ($F(2,64)=1,29, p<.28$), alors que cet effet est significatif pour le groupe plage ($F(2,64)=8,06, p<.0007$).

Nous obtenons la même différence entre ville et plage dans la deuxième expérience :



L'effet de la congruence ne provoque aucune facilitation pour le groupe ville (contraste sur l'amorce $F(2,62)=0,13$, $p<.88$), alors que cet effet de facilitation est significatif pour le groupe plage ($F(2,62)=8,26$, $p<.0006$).

Il est étonnant qu'il n'y ait d'effet de la congruence seulement pour le groupe plage. Cela signifie-t-il que les interférences globales et locales dépendent des catégories de scènes naturelles ? Nous proposons deux hypothèses pour expliquer ces différences intergroupes.

Tout d'abord, dans les deux expériences, la catégorisation d'une plage était plus lente que celle d'une ville. Dans l'expérience 1, cette différence n'est pas significative, $F(1,32)=3,43$, $p<.07$, mais sensible, puisque la moyenne des temps de catégorisation d'une plage est de 424ms, contre 391ms pour les villes. Dans l'expérience 2, cette différence est significative, $F(1,31)=13,88$, $p<.0008$, avec une moyenne des temps de catégorisation de 460ms pour les plages et de 371ms pour les villes. Les temps de réponse étant plus rapides pour catégoriser une ville, il est possible que les traitements mis en jeu dans cette reconnaissance s'affranchissent de l'aide apportée par l'amorce par une stratégie différente de celle de la reconnaissance de plage.

Une autre hypothèse peut expliquer à la fois les temps de catégorisation plus rapides observés pour les villes ainsi que la moindre influence de la congruence entre amorce et cible. Dans le masque que nous utilisons, les quatre premières images (voir annexe B) ont des SA équilibrés entre les différentes orientations. Cependant, les images 5 et 6 ont des SA présentant de fortes amplitudes selon les orientations horizontales, qui sont typiques des scènes de ville. Les images 7 et 8 exhibent également, dans une moindre mesure, des orientations horizontales, et presque aucune orientations

verticales. Ainsi, la neutralité du masque pourrait être mise en cause, puisqu'il pourrait, d'après le modèle décrit en première partie, 'amorcer' la catégorisation des villes.

L'hypothèse du masque ne peut cependant pas tout expliquer puisque ce biais est induit pas le fenêtrage de Butterworth que j'avais jugé logique d'appliquer aux images du masques. A l'origine, les images du masque non fenêtrées ne présentaient pas ce biais : les orientations dans les SA des images du masque étaient équitablement distribuées, et ne favorisaient pas la perception de l'une ou l'autre des catégories par le biais du SA. Ce masque non fenêtré, et donc sans les biais que j'ai indiqué plus haut, a été utilisé dans d'autres expériences de catégorisation de villes et de plages, et l'on retrouvait les mêmes caractéristiques du groupe ville qu'ici (vitesse de catégorisation plus rapide et effet de la congruence moindre), bien que moins prononcées.

Il est probable que les deux hypothèses avancées précédemment participent conjointement à l'explication de l'absence d'effet de la congruence pour le groupe ville.

b) Interprétations des résultats

Nous avons voulu tester les interactions entre la perception globale et la perception locale dans les scènes visuelles. Toutes les précautions que nous avons prises visent à rendre compte de ces interactions dans la perception d'une même scène naturelle. Ainsi, les images locales étaient extraites d'images globales (validité 'locale' des imagerie), et la dissociation entre local et global dans le protocole d'amorçage nous a permis de contrôler la partie locale de l'image qui était traitée relativement à l'image globale perçue.

Deux objections peuvent cependant être levées quant à la validité de ce protocole dans la simulation d'interactions globales/locales sur un même stimulus. La première est que les imagerie n'appartenaient pas à l'image globale à laquelle elles étaient associées, et la seconde est que le protocole d'amorçage induit un biais en dissociant temporellement la perception du local et du global.

Pour la première objection, d'après le modèle présenté en première partie, on peut justifier le fait que l'imagerie ne soit pas extraite de l'image globale à laquelle elle est associée par le fait que dans la perception implicite du traitement du stimulus, c'est le spectre d'amplitude qui déterminerait l'information relative à la catégorie. Donc les interactions entre catégorie locale et catégorie globale devraient être du même ordre si l'on ne considère, pour le choix des imagerie, que l'information du SA et non notre reconnaissance explicite (voir première partie, section 3c). Un moyen d'appuyer fortement cette hypothèse dans le cadre de notre étude serait de reproduire exactement les deux expériences que nous avons conduites, mais avec des images 'hybrides' en amorce, en suivant le même principe que l'expérience présentée dans la première partie, section 3c. Si nous obtenions un effet de la congruence des SA avec des résultats similaires à ceux obtenus ici, et qu'il n'y avait pas d'effet de la congruence des phase (le terme de congruence se réfère ici à la définition donnée pour l'expérience sur les images hybrides), alors cela montrerait que c'est réellement l'information du SA qu'il nous importe de traiter dans les interactions globales/locales.

Pour la deuxième objection, on doit en effet reconnaître que l'hypothèse qu'un paradigme d'amorçage rend compte des interactions global/local de la même manière que si l'on étudiait ces interactions dans un même stimulus est matière à discussions. Pour le savoir, la réalisation de l'expérience

suivante pourrait être intéressante. En prenant l'exemple des expériences sur les stimuli hiérarchiques à base de lettres, notre procédure conduirait à tester, pour l'influence du local sur le global, la congruence entre amorce et cible avec en amorce une petite lettre et en cible une grande lettre composée des mêmes petites lettres que l'amorce. La condition de congruence serait le lien entre les petites lettres et la grande lettre. Qu'obtiendrions-nous comme résultats ? Si dans une telle expérience nous obtenions un effet de la congruence, cela serait contraire aux résultats obtenus dans les conditions expérimentales classiques de stimuli hiérarchiques. Cependant, l'interprétation de ces résultats serait délicates, étant donné que sous certaines conditions on peut obtenir un effet de précédence et d'interférence locale (stimuli hiérarchiques dilatés, filtrage des basses fréquences) [HUG96]. Mais étant donné qu'il n'est pas possible de montrer l'équivalence entre le protocole utilisé et ce qui se passe lors du traitement d'un stimulus visuel, l'hypothèse que ce protocole peut généraliser aux interactions existant au sein d'un même stimulus doit être avancée avec beaucoup de précautions.

La question de savoir si l'objectif initial du protocole peut être remis en cause n'invalide cependant pas les effets de facilitations perceptives observés entre échelle globale et échelle locale, mais rend leur interprétation différente. Ainsi, si l'on considère que ce protocole ne teste pas les interactions global/local dans de la perception d'un stimulus, nous avons toutefois montré l'existence d'une influence de la perception globale sur la perception locale, et de la perception locale sur la perception globale, avec des stimuli de scènes naturelles. Cette influence dépend du type de congruence des catégories d'appartenance et des SA des images globales et locales. Si catégories d'appartenances et SA de l'amorce et de la cible sont congruents, on observe une facilitation de la catégorisation de la cible par rapport à la condition où ils ne sont pas congruents.

Conclusion

Plusieurs facteurs sont à considérer pour conclure sur les deux expériences que nous avons présentées dans cette étude. Ces expériences constituaient une première étape dans l'investigation des interactions entre propriétés globales et propriétés locales des scènes naturelles, et se basaient sur deux hypothèses :

- Le spectre d'amplitude est un vecteur d'information clef dans les processus implicites de catégorisation de scène par notre système visuel.
- Le protocole expérimental d'amorçage utilisé permet l'étude des interactions entre propriétés globales et propriétés locales des scènes naturelles.

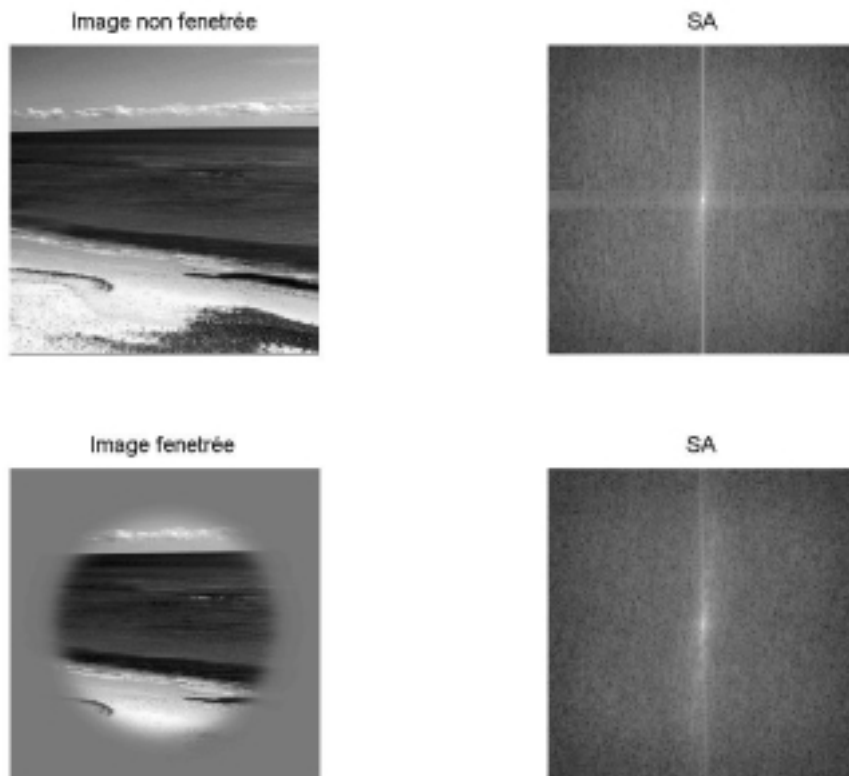
Les résultats obtenus ont montré qu'il y avait un effet de la perception globale sur la perception locale, ainsi que de la perception locale sur la perception globale. Cependant, ces expériences étant princeps en la matière, elles ne permettent ni de conclure que ces interactions perceptives entre global et local se produisent du fait de la similarité / dissimilarité des spectres d'amplitudes manipulés conjointement au catégories de scènes, ni qu'elles se produiraient de la même manière si elles n'étaient pas dissociées temporellement par le paradigme d'amorçage.

Afin de valider le rôle du spectre d'amplitude dans les interactions observées, il est nécessaire de conduire à nouveau les expériences menées, mais en remplaçant les stimuli d'amorce par des images 'hybrides' (phase et spectre d'amplitude manipulés séparément). Quant à la validité du protocole concernant la généralisation aux interactions global/local pouvant exister au sein d'un même stimulus, il n'est pas possible de conclure avec les données dont nous disposons actuellement, et le principe de précaution suggérerait de s'abstenir de telles conclusions.

Annexe A : Fenêtrage de Butterworth

Intérêt

Les discontinuités en bordure d'image se traduisent lors de la transformée de Fourier discrète par l'apparition de sinus cardinaux selon les orientations verticales et horizontales, comme illustré ci-après :



Exemple d'apparition de sinus cardinaux dans le cas d'une image non fenêtrée

L'apodisation de l'image est une opération qui consiste à enlever ces discontinuités en bordure d'image en ramenant progressivement à une même valeur tous les points en bordure de l'image. Le fenêtrage de Butterworth est un moyen d'effectuer cette apodisation.

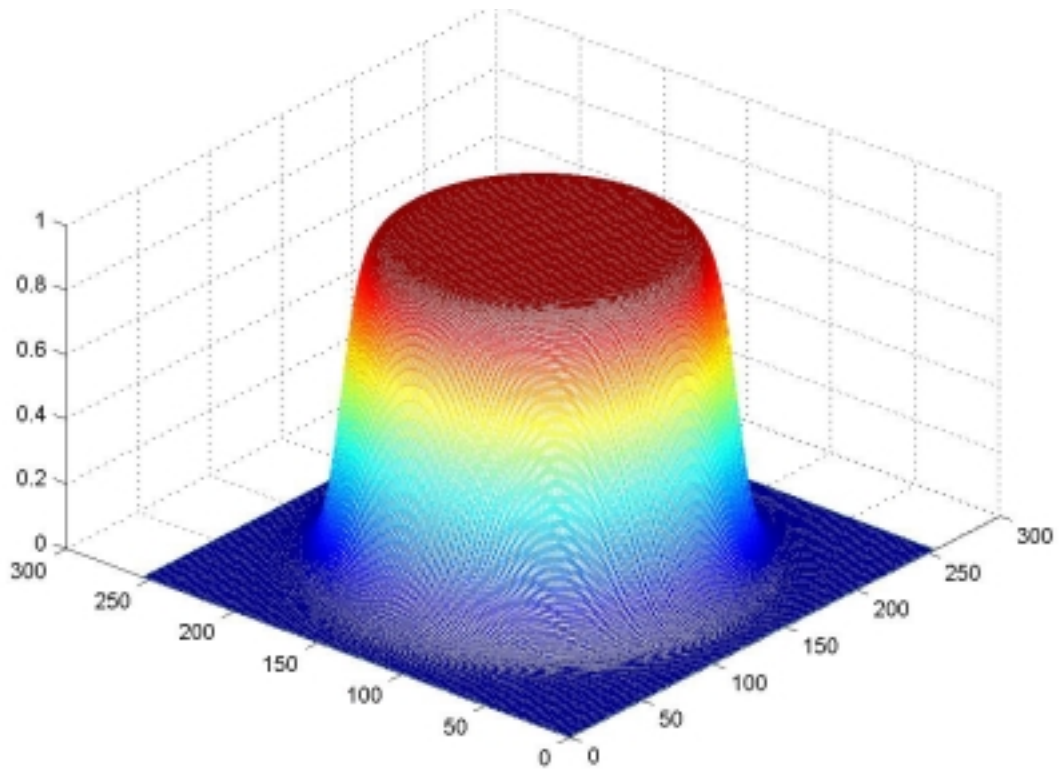
Procédé

L'obtention de nos images fenêtrées est conduite de la manière suivante :

- Soit $i(x,y)$ l'image à fenêtrer, $x,y \in [1,N]$ où N est la taille d'un côté de l'image.

- Soit
$$h(x,y) = \frac{1}{1 + \left(\sqrt{\left(\frac{x-N/2}{N} \right)^2 + \left(\frac{y-N/2}{N} \right)^2} / f_{coup} \right)^{ordre}}$$
 la fonction

d'apodisation de Butterworth. Les paramètres f_{coup} et $ordre$ ont été fixé de façon empirique à respectivement 0.35 et 20. Cette fonction est représentée ci-dessous pour $N=256$:

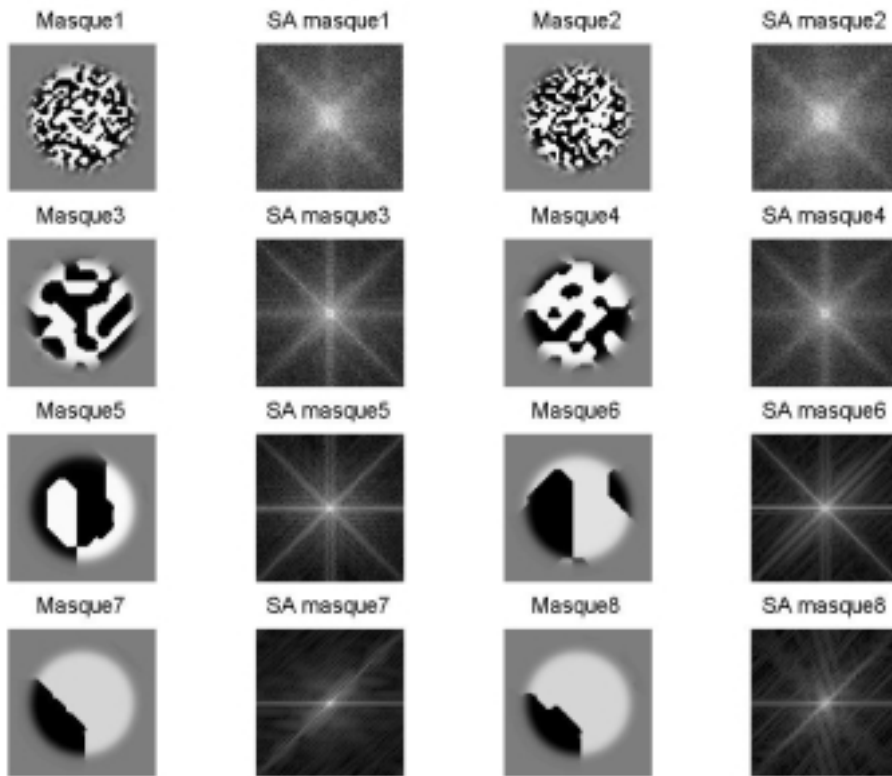


Fonction d'apodisation de Butterworth

- Soit $g(x,y)=128$ une fonction constante d'une luminosité moyenne (les images sont en 256 niveaux de gris)
- L'image fenêtrée $f(x,y)$ est obtenue par : $f(x,y)=i(x,y)h(x,y)+g(x,y)(1-h(x,y))$

Annexe B : Masque dynamique

Images et spectre d'amplitude des huit images utilisées pour le masque



Annexe C : Images choisies

Statistiques sur les images

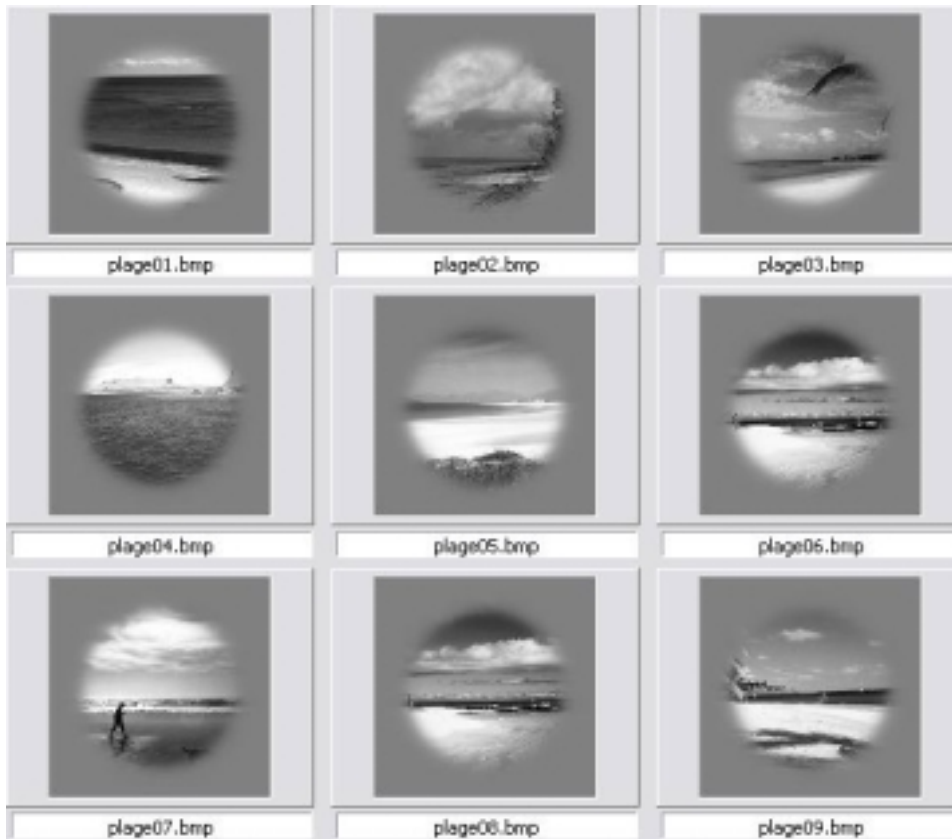
Images globales

Catégorie	Nom	Moyenne	Variance (10 ³)	Energie image (10 ⁹)	Energie spectre (10 ¹²)
Plage	Plage01	117.9630	1.7291	1.0253	7.4265
Plage	Plage02	130.2260	1.1806	1.1652	5.2244
Plage	Plage03	131.3570	1.3286	1.1838	5.7729
Plage	Plage04	136.0035	1.8824	1.3356	8.0847
Plage	Plage05	145.6295	1.5579	1.4920	6.6909
Plage	Plage06	137.4477	1.8171	1.3572	7.8041
Plage	Plage07	142.0201	1.7699	1.4378	7.6016
Plage	Plage08	135.5968	1.3982	1.2966	6.0052
Plage	Plage09	137.4253	1.4345	1.3317	6.1611
Ville	Ville01	118.3340	2.4070	1.1017	9.956
Ville	Ville02	130.4263	1.7282	1.2281	7.423
Ville	Ville03	131.3604	2.2409	1.2777	9.625
Ville	Ville04	140.9355	1.8056	1.4201	7.755
Ville	Ville05	118.1565	1.5762	1.0182	6.770
Ville	Ville06	132.1132	1.2964	1.2288	5.568
Ville	Ville07	129.0487	1.7010	1.2029	7.306
Ville	Ville08	125.8246	1.9781	1.1672	8.496
Ville	Ville09	125.7143	1.9770	1.1653	8.491

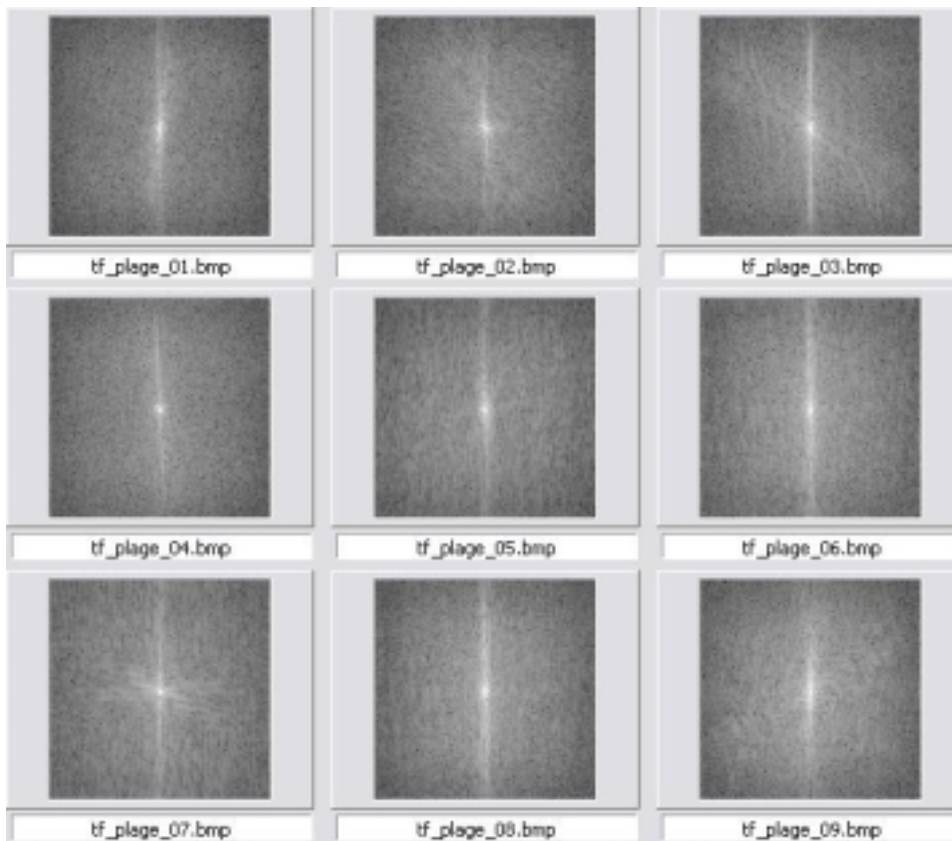
Images locales

Catégorie	Nom	Moyenne	Variance	Energie image (10 ⁹)	Energie spectre (10 ¹²)
Plage	L_Plage01	127.1821	133.5212	1.0688	0.5735
Plage	L_Plage02	128.1624	211.2928	1.0903	0.9075
Plage	L_Plage03	127.3518	180.7675	1.0747	0.7764
Plage	L_Plage04	127.9658	149.9757	1.0830	0.6441
Plage	L_Plage05	126.9536	134.8617	1.0651	0.5792
Plage	L_Plage06	128.2686	240.1589	1.0940	1.0315
Plage	L_Plage07	127.5038	186.7591	1.0777	0.8021
Plage	L_Plage08	128.7806	304.8890	1.1069	1.3095
Plage	L_Plage09	127.2248	162.7169	1.0714	0.6989
Ville	L_Ville01	127.1021	212.0031	1.0726	0.9105
Ville	L_Ville02	127.4534	102.7628	1.0711	0.4285
Ville	L_Ville03	127.8009	110.7313	1.0768	0.4797
Ville	L_Ville04	128.4770	213.6289	1.0958	0.9175
Ville	L_Ville05	127.3716	150.4641	1.0731	0.6462
Ville	L_Ville06	128.1436	153.7491	1.0862	0.6603
Ville	L_Ville07	128.6465	246.1612	1.1007	1.0572
Ville	L_Ville08	128.4862	145.4866	1.0914	0.6249
Ville	L_Ville09	128.4301	105.2724	1.0879	0.4521

Images globales de plages



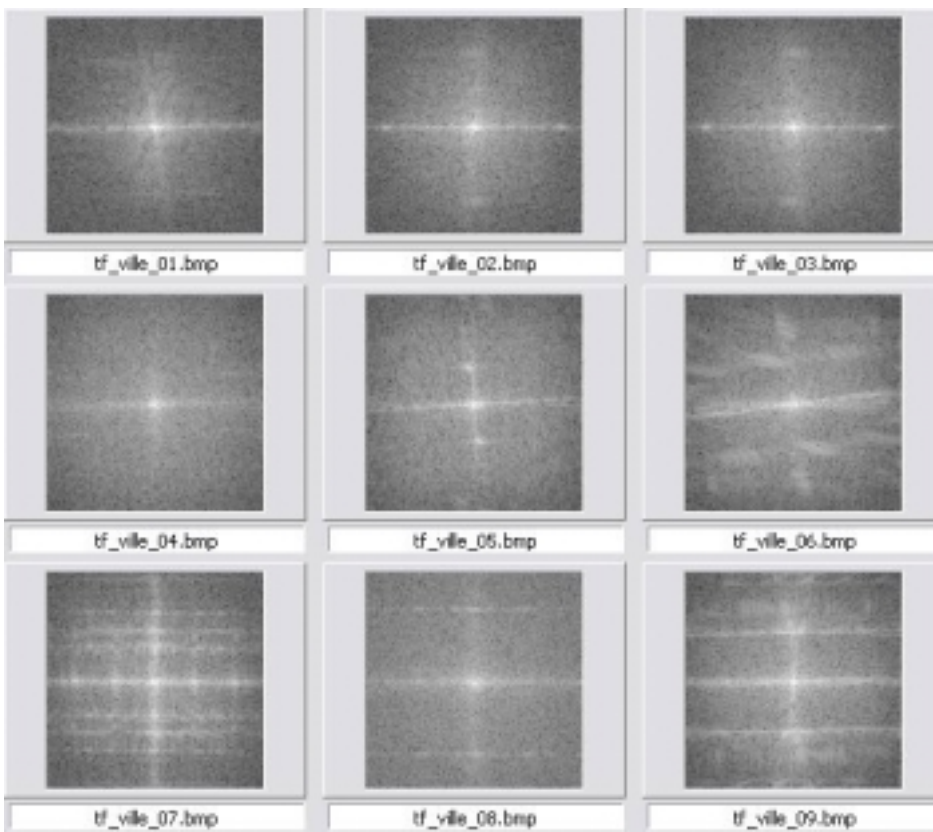
Log du spectre d'amplitude des images globales de plages



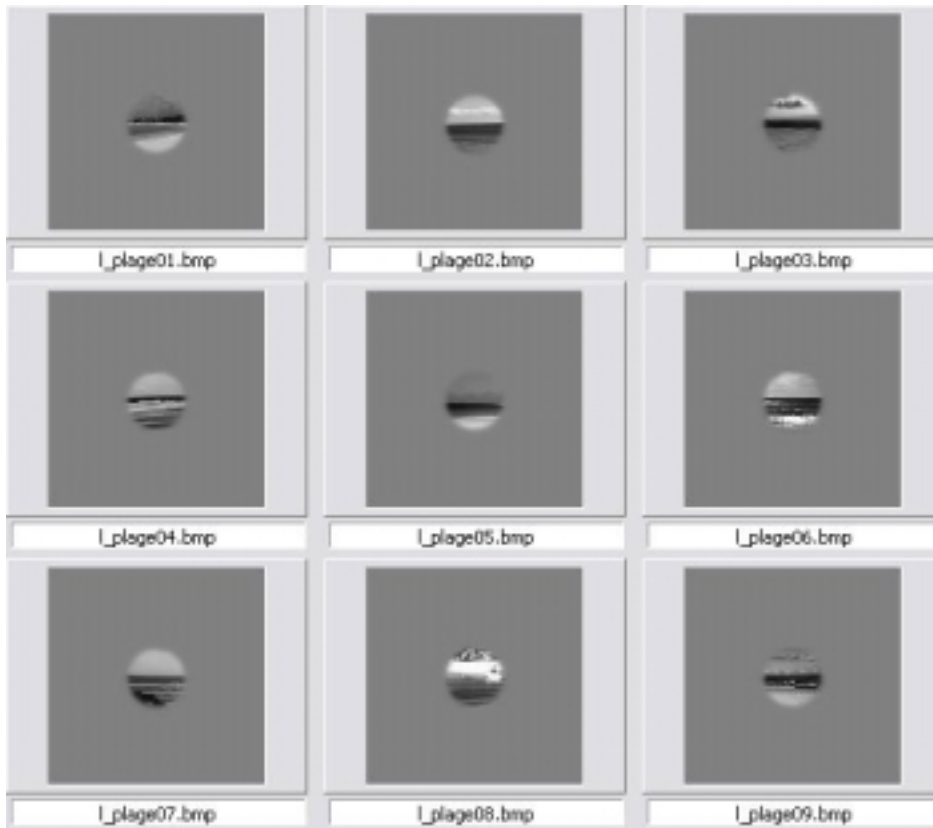
Images globales de villes



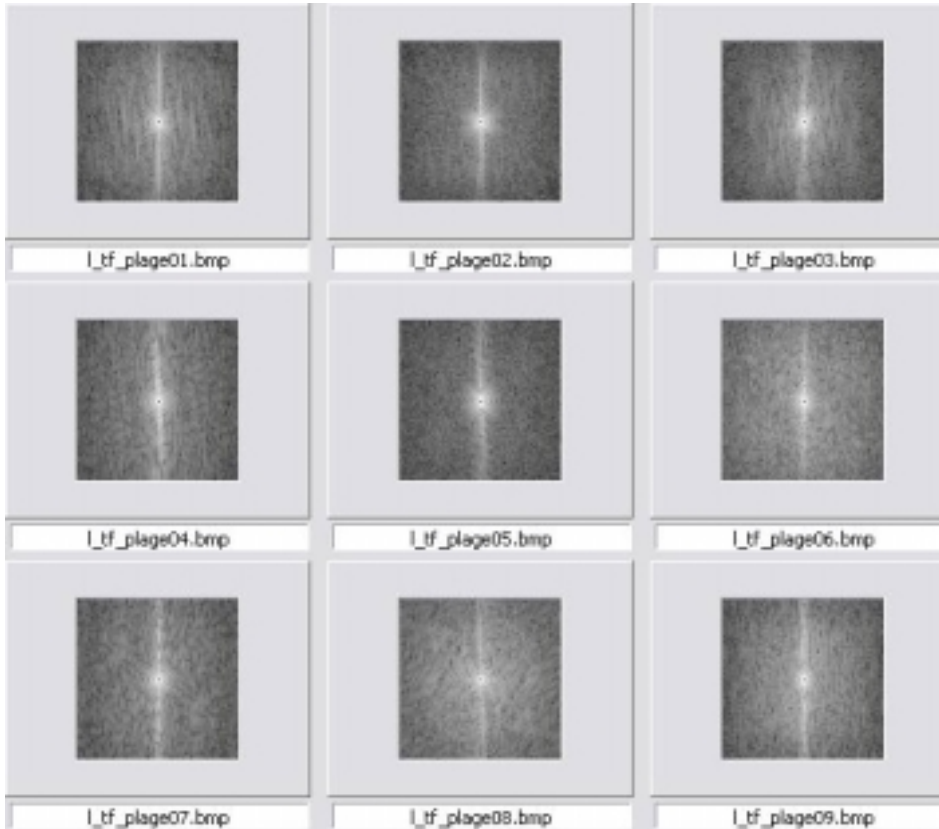
Log du spectre d'amplitude des images globales de villes



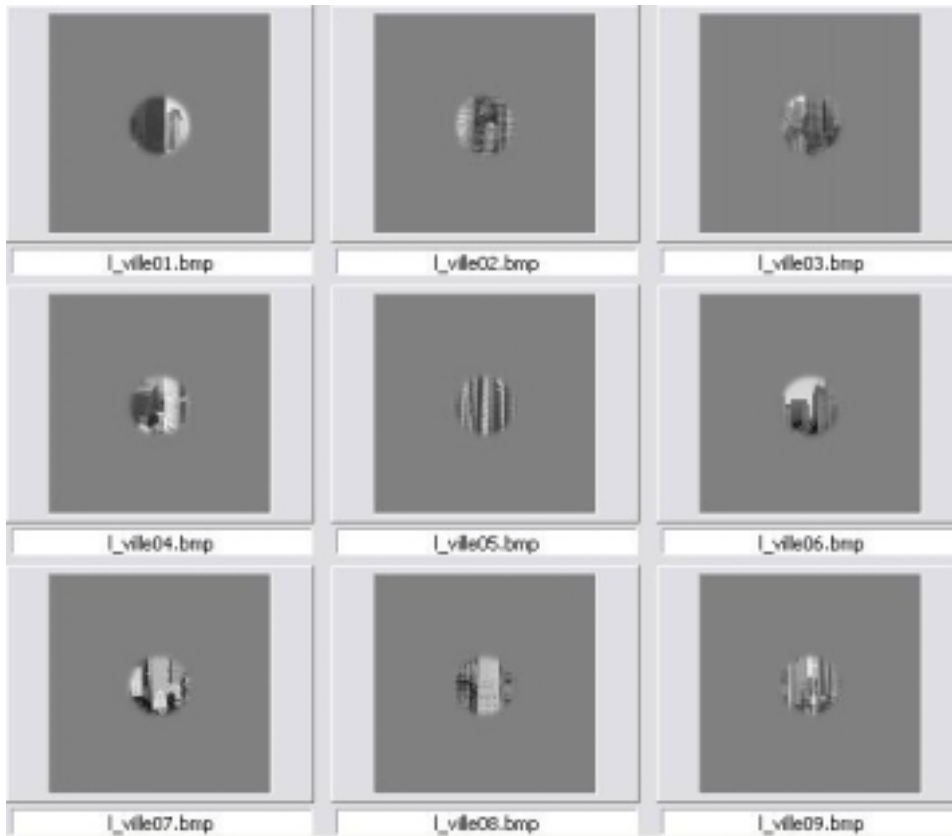
Images locales de plages



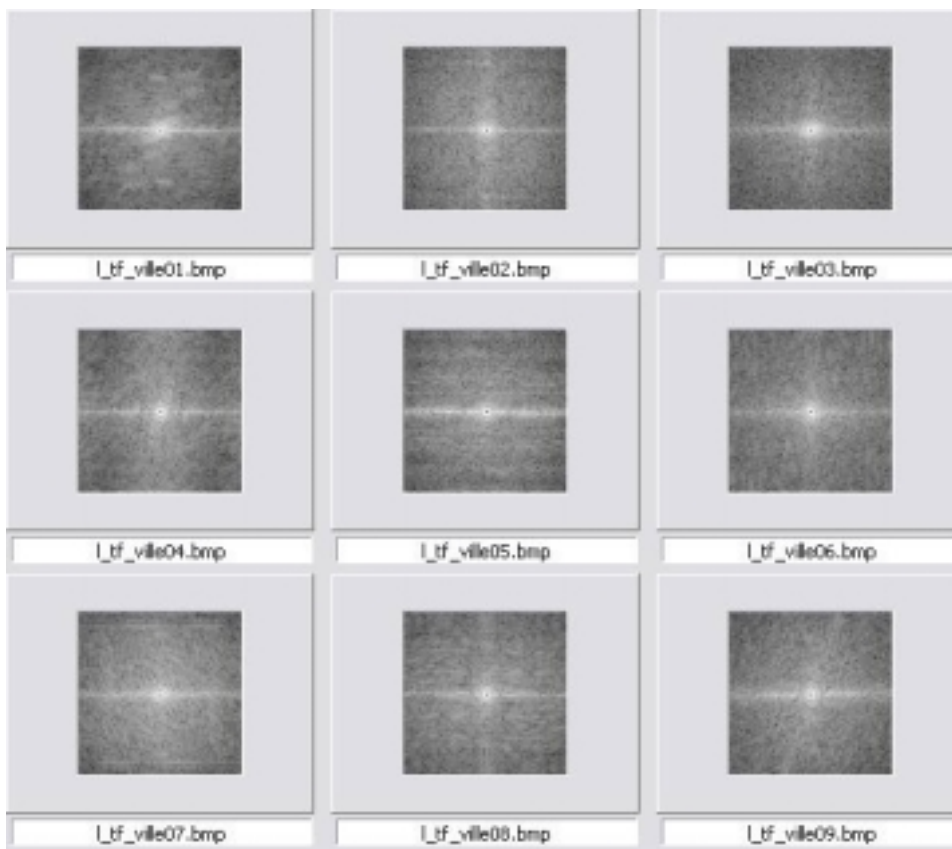
Log du spectre d'amplitude des images locales de plage



Images locales de villes



Log du spectre d'amplitude des images locales de villes



Annexe D : Ecrémage et distribution

Les temps de réaction des sujets suivent une distribution de type poissonnienne, et nous nous sommes basés pour l'élimination des valeurs jugées aberrantes sur deux critères :

- Tous les temps de réactions inférieurs à 200ms (C'est une limite inférieure du temps de réponse moteur pour réagir à un stimulus).
- Les temps de réactions qui sortaient trop largement de la distribution pour chacun des groupes dans chacune des deux expériences.

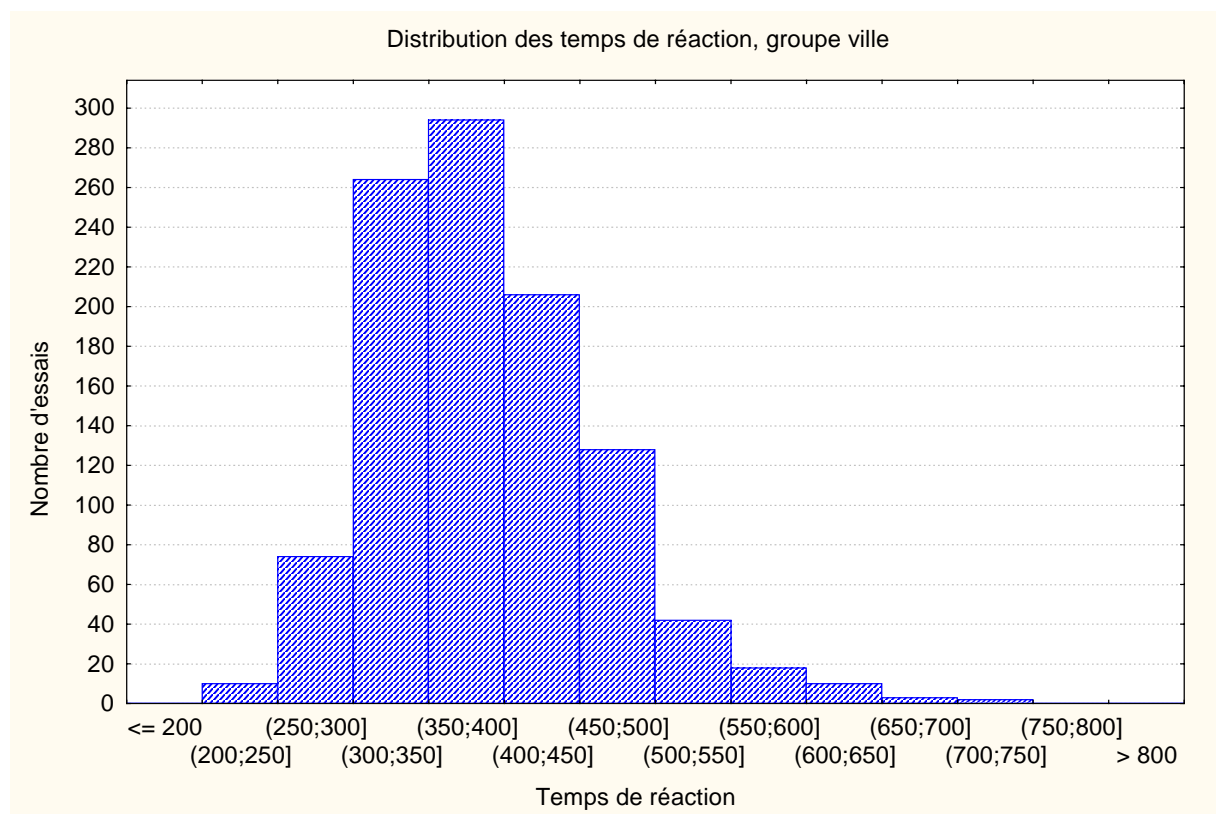
Expérience 1

Groupe ville

Temps de réaction supprimés

Catégorie de l'amorce	Catégorie de la cible	Temps de réaction (ms)
Ville	Ville	106
Plage	Ville	119
Neutre	Ville	159
Neutre	Ville	899
Ville	Ville	1204
Ville	Ville	1262
Neutre	Ville	2199

Distribution des temps de réactions

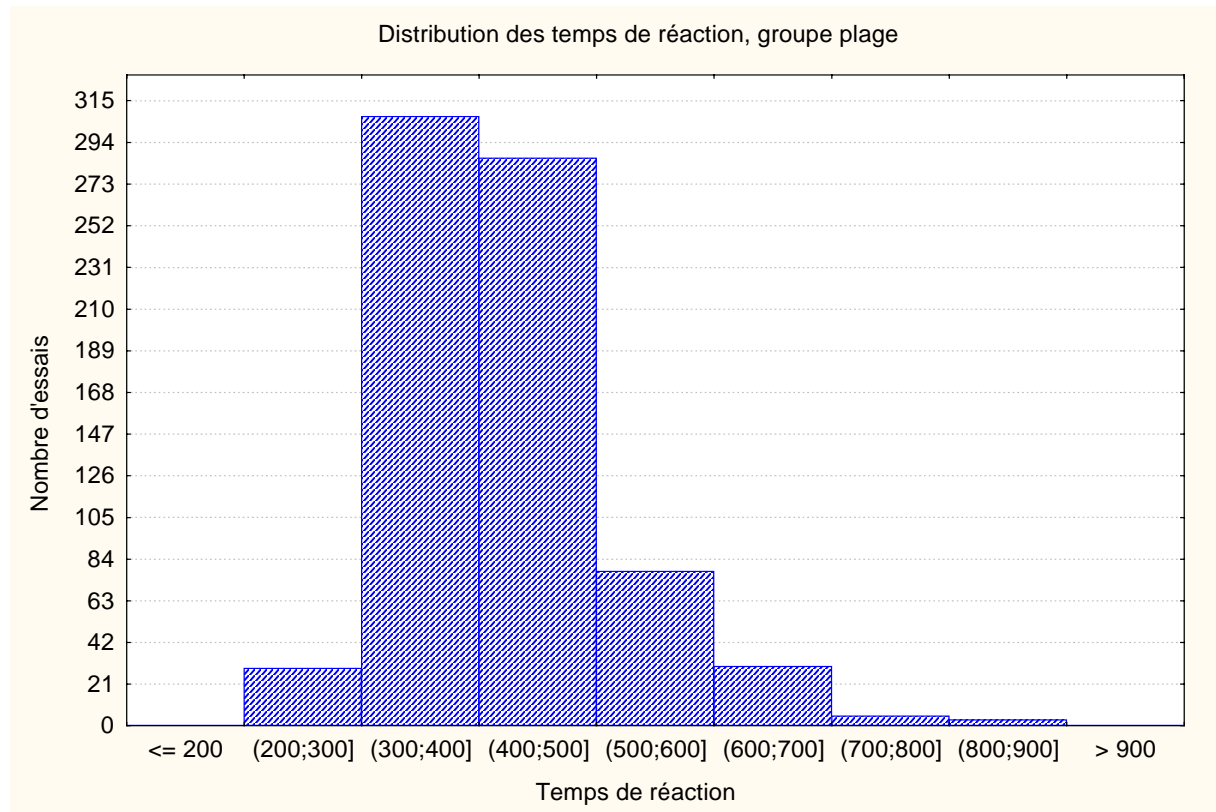


Groupe plage

Temps de réaction supprimés

Catégorie de l'amorce	Catégorie de la cible	Temps de réaction (ms)
Neutre	Plage	106
Ville	Plage	2199

Distribution des temps de réaction



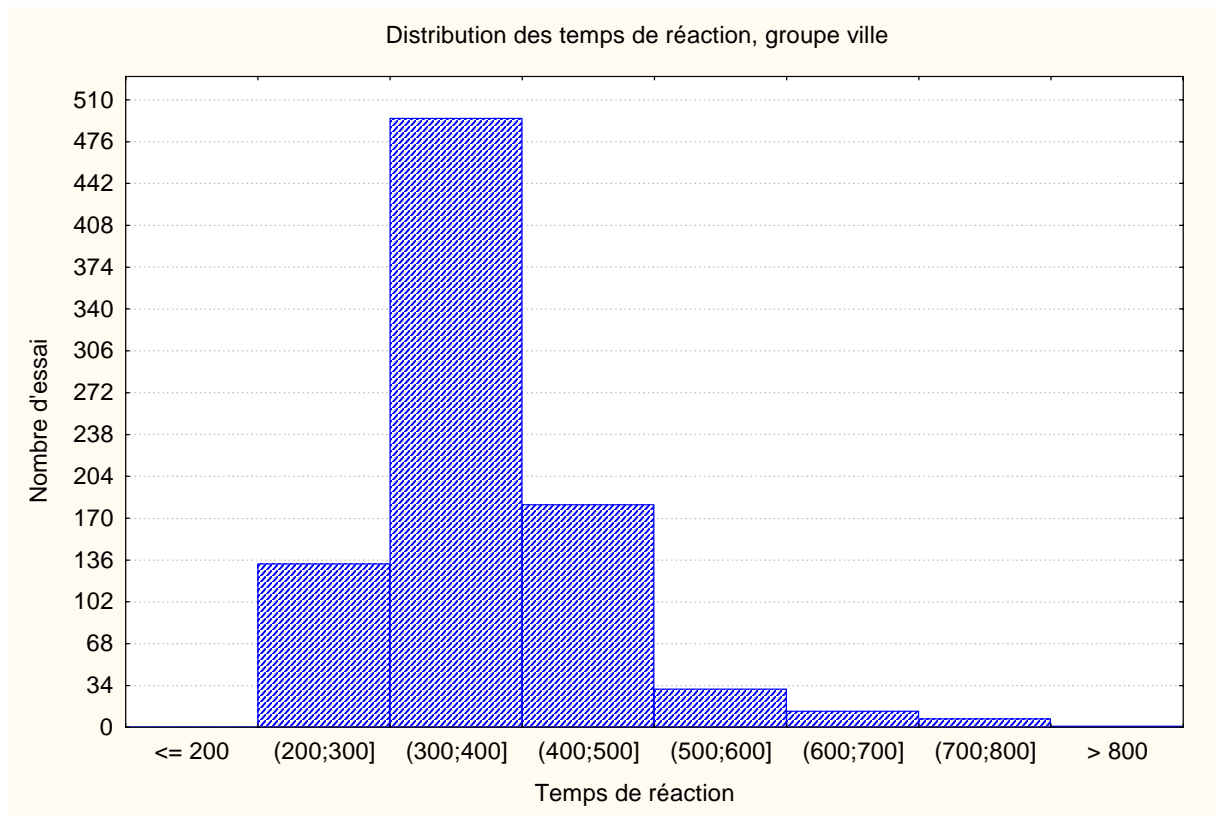
Expérience 2

Groupe ville

Temps de réaction supprimés

Catégorie de l'amorce	Catégorie de la cible	Temps de réaction (ms)
Neutre	Ville	50
Neutre	Ville	141

Distribution des temps de réaction



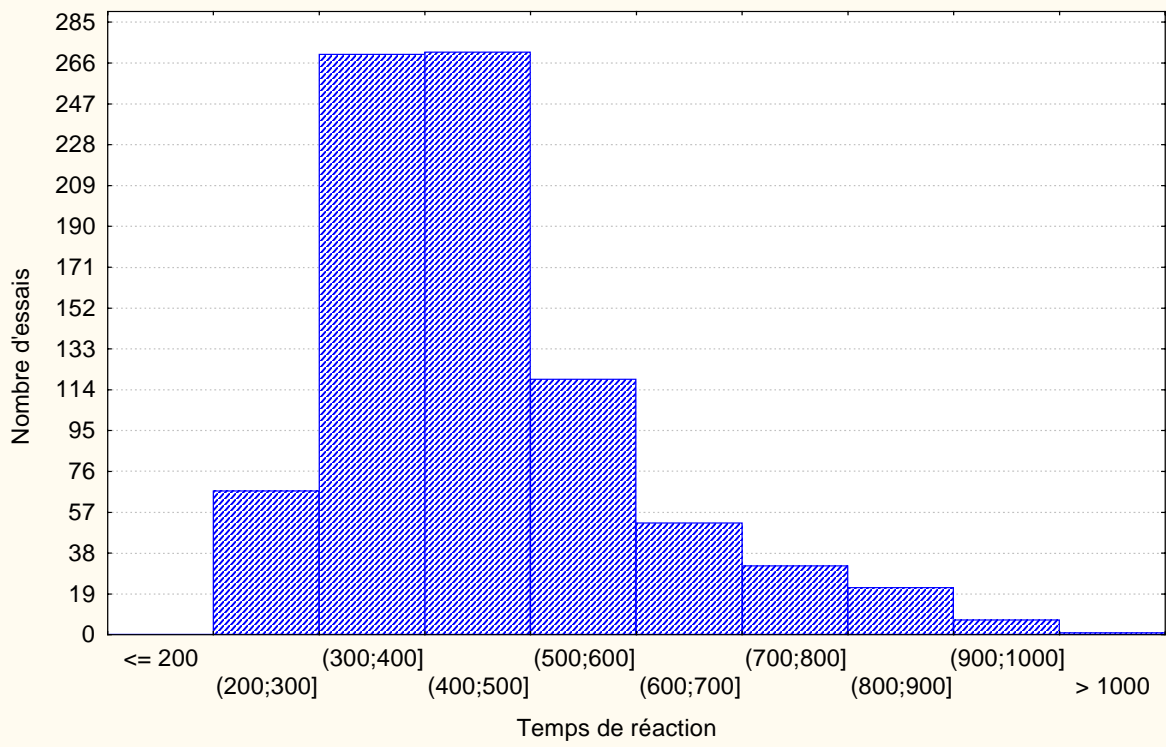
Groupe plage

Temps de réaction supprimés

Catégorie de l'amorce	Catégorie de la cible	Temps de réaction (ms)
Neutre	Plage	108
Neutre	Plage	140
Plage	Plage	142
Neutre	Plage	154
Neutre	Plage	1117
Ville	Plage	1125
Ville	Plage	1209
Plage	Plage	1425
Ville	Plage	1446
Ville	Plage	1504

Distribution des temps de réaction

Distribution des temps de réaction



Annexe E : Résultats et statistiques

Expérience 1

Temps de réaction moyen (en ms) par sujet et par condition

Sujet	Groupe	Session 1			Session 2		
		Neutre	Non congruent	Congruent	Neutre	Non congruent	Congruent
1	Ville	455	422	410	436	407	369
2	Ville	392	371	388	347	327	339
3	Ville	415	414	426	408	394	383
4	Ville	343	330	343	339	352	346
5	Ville	349	359	399	340	361	350
6	Ville	469	485	454	401	430	450
7	Ville	495	487	504	419	426	386
8	Ville	406	458	402	409	387	372
9	Ville	512	462	464	424	483	473
10	Ville	406	349	363	331	343	360
11	Ville	487	516	439	442	464	396
12	Ville	351	341	321	342	361	344
13	Ville	443	497	426	471	444	444
14	Ville	357	342	392	383	378	374
15	Ville	303	305	308	299	309	297
16	Ville	410	416	420	400	398	373
17	Ville	396	364	384	355	377	354
18	Ville	293	328	336	321	316	281
19	Ville	376	380	341	397	362	362
20	Ville	410	397	435	392	403	395
21	Plage	426	422	411	472	476	438
22	Plage	460	408	390	345	354	326
23	Plage	384	420	464	429	435	418
24	Plage	417	394	352	352	358	351
25	Plage	391	519	470	418	583	483
26	Plage	363	383	379	393	424	404
27	Plage	534	574	552	470	504	489
28	Plage	330	372	330	329	327	317
29	Plage	436	415	390	392	373	375
30	Plage	430	466	420	472	476	480
31	Plage	432	462	398	420	423	410
32	Plage	533	600	444	488	470	450
33	Plage	483	486	471	352	378	370
34	Plage	393	360	378	358	435	381

ANOVA : Analyse canonique des temps de réponses

	dl effet	CM effet	dl erreur	CM erreur	F	niveau p
G	1	52242,6797	32	15188,2617	3,43	0,072
S	1	17906,1504	32	1923,45837	9,31	** 0,0045
A	2	5304,47314	64	706,311584	7,51	** 0,0012
G-S	1	84,2671585	32	1923,45837	0,044	0,83
G-A	2	2142,71143	64	706,311584	3,03	0,055
S-A	2	103,677544	64	492,250946	0,21	0,81
G-S-A	2	413,562469	64	492,250946	0,84	0,44

ANOVA : Analyse canonique des pourcentage d'erreurs

	dl effet	CM effet	dl erreur	CM erreur	F	niveau p
G	1	0,0027	32	0,00067	4,002	0,054
S	1	0,0004	32	0,00036	1,301	0,26
A	2	0,0011	64	0,000236	4,65	** 0,013
G-S	1	0	32	0,00036	0	1
G-A	2	0,001	64	0,00023	4,281	** 0,018
S-A	2	0,00001	64	0,00013	0,145	0,86
G-S-A	2	0,0002	64	0,00013	2,17	0,12

Expérience 2

Temps de réaction moyen (en ms) par sujet et par condition

Sujet	Groupe	Session 1			Session 2		
		Neutre	Non congruent	Congruent	Neutre	Non congruent	Congruent
1	Ville	320	285	323	298	283	269
2	Ville	450	440	409	410	400	403
3	Ville	344	334	335	332	343	367
4	Ville	353	318	325	374	341	354
5	Ville	415	408	379	315	321	336
6	Ville	321	385	290	363	326	351
7	Ville	290	318	313	377	366	387
8	Ville	401	426	369	357	376	363
9	Ville	401	368	409	390	405	439
10	Ville	357	360	350	352	360	382
11	Ville	372	344	342	326	367	358
12	Ville	399	420	434	378	400	332
13	Ville	407	403	396	384	381	398
14	Ville	377	360	334	323	374	351
15	Ville	458	425	462	375	333	344
16	Ville	408	457	466	482	503	445
17	Plage	612	609	640	500	610	460
18	Plage	754	749	707	449	498	478
19	Plage	450	500	495	402	414	358
20	Plage	427	430	455	405	370	374
21	Plage	479	434	412	374	406	394
22	Plage	506	448	532	547	524	462
23	Plage	526	489	506	508	523	551
24	Plage	331	436	358	287	352	297
25	Plage	359	457	324	351	394	363
26	Plage	514	571	538	444	396	387
27	Plage	665	734	634	465	514	412
28	Plage	512	505	429	354	391	328
29	Plage	586	510	416	604	523	587
30	Plage	419	427	405	358	365	350
31	Plage	316	342	340	331	333	351
32	Plage	344	387	324	375	452	330
33	Plage	633	594	675	439	467	453

ANOVA : Analyse canonique des temps de réponse

	dl effet	CM effet	dl erreur	CM erreur	F	niveau p
G	1	386708,688	31	27857,9531	13,88	** 0,00078
S	1	79262,7656	31	7716,9209	10,27	** 0,0031
A	2	4723,76807	62	940,244995	5,02	** 0,0095
G-S	1	45191,0859	31	7716,9209	5,85	** 0,022
G-A	2	2929,46802	62	940,244995	3,12	0,051
S-A	2	202,708893	62	909,125488	0,22	0,80
G-S-A	2	307,415588	62	909,125488	0,34	0,71

ANOVA : analyse canonique des taux d'erreurs

	dl effet	CM effet	dl erreur	CM erreur	F	niveau p
G	1	0,00257568	31	0,00139593	1,85	0,18
S	1	0,00654718	31	0,00062619	10,4	** 0,0029
A	2	0,0017655	62	0,00059132	2,98	0,058
G-S	1	0,00045055	31	0,00062619	0,72	0,40
G-A	2	0,00053578	62	0,00059132	0,91	0,41
S-A	2	7,8761E-05	62	0,00028458	0,28	0,76
G-A-S	2	0,00012379	62	0,00028458	0,44	0,65

Bibliographie

- [BAC02] Bacon N., Thorpe S. ; *Catégorisation visuelle ultra-rapide et masquage rétroactif ; Rapport DEA, CERCO Toulouse, 2002.*
- [CNGH02] Chauvin, A., Guyader, N., Marendaz, C., & Hérault, J. ; *Argument for scene categorisation with image amplitude spectra. Perception, 31(ECVP 2002), 132-133*
- [GUY02] Guyader N., Hérault J.; *Représentation espace-fréquences pour la catégorisation d'images*
- [HEN99] Henderson J. , Hollingworth A. ; *high level scene perception ; Annual review psychology 1999 p243-271.*
- [HUG96] Hughes H, Nozawa G, Kitterle F. ; *Global precedence, spatial frequency channels, and the statistics of natural images ; Journal of cognitive neuroscience, 1996 volume 8 n°3, p197-230*
- [LEE96] Lee T. ; *Image representation using 2D gabor wavelets ; IEEE transactions on pattern and machine intelligence, volume 18 n°10 october 96.*
- [OLI94] Oliva A. ; *Perception de scènes ; Thèse institut national polytechnique de Grenoble 1994.*
- [OLI01] Oliva A., Torralba A ; *Modeling the shape of the scene : A holistic representation of the spatial envelope ; International journal of computer vision 2001, 43(3), p145-175.*