



HAL
open science

De la prise d'information visuelle à la formation d'impressions : apports de l'oculométrie pour l'étude des processus de la perception et de la cognition visuelle des objets manufacturés

Thomas Couronné

► **To cite this version:**

Thomas Couronné. De la prise d'information visuelle à la formation d'impressions : apports de l'oculométrie pour l'étude des processus de la perception et de la cognition visuelle des objets manufacturés. Interface homme-machine [cs.HC]. Université Joseph-Fourier - Grenoble I, 2007. Français. NNT : . tel-00292218

HAL Id: tel-00292218

<https://theses.hal.science/tel-00292218v1>

Submitted on 30 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE GRENOBLE I – JOSEPH FOURIER
Laboratoire de l'Image, de la Parole, du Signal et de l'Automatique
CNRS UMR 5216

Thèse en vue de l'obtention du grade de

Docteur de l'Université de Grenoble I

Spécialité :

Ingénierie de la cognition, de l'interaction, de l'apprentissage et de la création

Présentée par **Thomas Couronné**

« De la prise d'information visuelle à la formation d'impressions : apports de l'oculométrie pour l'étude des processus de la perception et de la cognition visuelle des objets manufacturés »

Sous la direction académique de

Michel Dubois et Anne Guérin-Dugué

En collaboration avec le groupe PSA PEUGEOT CITROEN

Sous la tutelle industrielle de **Pauline Faye**

JURY :

Thierry Baccino, Professeur, Université de Nice-Sophia Antipolis (Rapporteur)

Peter De Graef, Professeur, Université de Leuven, Belgique (Rapporteur)

Michel Dubois, Professeur, Université Pierre Mendès-France, Grenoble

Pauline Faye, Ingénieur, Groupe PSA Peugeot Citroën, Vélizy

Anne Guérin-Dugué, Professeur, Université Joseph Fourier, Grenoble

Christian Marendaz, Professeur, Université Pierre Mendès-France, Grenoble

Jean-Christophe Popieul, Professeur, Université de Valenciennes et du Hainaut-Cambrésis

Thèse soutenue publiquement à l'Institut National Polytechnique de Grenoble le 18 Décembre 2007

Résumé /Abstract

Mots-clés : traitement de données, oculométrie, mouvements des yeux, perception visuelle, cognition visuelle, évaluation, jugement, produit manufacturé, automobile

L'intégration des facteurs humains dans la conception et le développement industriel est aujourd'hui un enjeu majeur de l'innovation technologique, et plus particulièrement pour l'automobile. L'étude des processus d'évaluation des produits manufacturés permet ainsi d'apporter des informations clés sur la manière dont le produit va être perçu puis jugé. Par son approche pluridisciplinaire à la confluence des sciences de l'ingénieur et des sciences cognitives, notre travail contribue à la compréhension des processus de la cognition visuelle et de formation d'impression par l'usage de l'oculométrie. Ainsi, les comportements oculomoteurs mesurés en situation d'évaluation visuelle du produit automobile vont être caractérisés et catégorisés par des méthodes statistiques empruntées au traitement des images, à la biologie, à l'agronomie ou la psychophysique. Les liens entre l'orientation du regard, l'attention visuelle et l'intégration cognitive des informations visuelles pour l'évaluation seront ensuite étudiés. Plusieurs paradigmes expérimentaux sont conçus : évaluation monadique ou par paire, à durée d'exposition libre ou limitée, depuis des positions d'observations libres ou contraintes. Il s'avère que les processus qui guident l'orientation du regard lors de l'évaluation visuelle de produits photographiés, tels que les habitacles automobiles, sont nombreux et imbriqués de manière complexe tant spatialement que temporellement. Ce travail développe non seulement un ensemble de protocoles et de préconisations d'usage de l'oculométrie, mais aussi une démarche d'analyse de ces données par des traitements en cascade, afin de mettre en évidence le traitement visuel des zones d'intérêt et les effets des consignes. Finalement, un ensemble de méthodes est proposé pour estimer les contributions relatives des différents processus attentionnels sur la structuration des distributions de fixations.

Key-words: data analysis, eye movement, visual perception, visual cognition, assessment, evaluation, judgment, manufactured objects, automotive

Today, Human factors are highly strategic in the industrial process of design and technical innovation, and have to be considered in particular for automotive industry. Especially, the consumer's behavior studies are very useful for a better understanding of a judgment process. In that way, our work deals with engineering and cognitive sciences. It contributes by the use of eye movements to analyze and model visual cognition and assessment process. At first, the eye movements are characterized and classified with statistical tools coming from image processing, biology, agronomy or psychophysics. Then, we observe links between gaze orientation, visual attention, and cognitive integration of visual information. Different experimental paradigms are built: monadic and pair-wise evaluation, free or fixed exposition duration, constrained or free point of view. In the context of automotives evaluations, gaze behavior seems to be influenced by a lot of factors, organized in a complex manner. That's why we suggest new protocols and recommendations for the use of eye-movement measures, in these complex situations. A "statistical eye movement analysis toolbox" and a "cascading treatment methodology" are built for observation of visual behavior on areas of interest and for visual information integration. Finally, we suggest new methods to evaluate the different attention process effects on eye movements.

Cette thèse de doctorat a été réalisée au sein de :

PSA Peugeot Citroën

Direction de la Recherche et de l'Innovation Automobile, Entité Perceptions
et Facteurs Humains,

Centre Technique de Vélizy,

Route de Gisy, 78943 Vélizy-Villacoublay

Sous la supervision du :

**LABoratoire Grenoblois de l'Image, de la Parole, du Signal et de
l'Automatique**

(GIPSA-LAB, ex-Laboratoire des Images et Signaux)

Institut National Polytechnique de Grenoble,

46 avenue Félix Viallet,

F-38031 Grenoble

Remerciements

Ce travail n'aurait pas été possible sans le support et l'accompagnement de nombreuses personnes, que je tiens à remercier ici.

Je souhaite tout d'abord exprimer mes chaleureux remerciements aux superviseurs de cette thèse pour leur disponibilité, leur expertise, et leur ouverture d'esprit dans ce travail pluridisciplinaire. Par ordre alphabétique, je remercie :

Michel Dubois (Université Pierre Mendès-France, Laboratoire de Psychologie Sociale, Grenoble) pour ses conseils et ses relectures attentives,

Pauline Faye (PSA PEUGEOT CITROEN) pour sa gestion dynamique du projet, sa motivation, et pour ses conseils en statistiques,

Anne Guérin-Dugué (Université Joseph-Fourrier, Gipsa-Lab, Grenoble) pour son suivi attentif, son excellent sens de la pédagogie, et pour son investissement permanent,

Christian Marendaz (Université Pierre Mendès-France, Laboratoire de Psychologie et Neuro-Cognition, Grenoble) pour toutes les idées et suggestions, ainsi que pour ses conseils ayant permis de bien positionner cette thèse et ses enjeux,

Brigitte Meillon (CLIPS-Laboratoire d'informatique de Grenoble) pour son expertise technique, sa gestion des petits détails qui simplifient la vie et pour sa sympathie.

Je tiens ensuite à remercier messieurs Thierry Baccino (Université de Nice), Peter De Graef (Université de Leuven, Belgique) et Jean-Christophe Popieul pour avoir accepté d'examiner ce présent travail de doctorat.

Mes remerciements sincères et cordiaux s'adressent à tous mes collègues et amis du Laboratoire PErceptions et Facteurs Humains (PSA PEUGEOT CITROEN), qui aussi bien « permanents », « prestataires » que « stagiaires », ont comptés pendant ces trois années.

En particulier, de chaleureux remerciements à Anne Bardot, qui a monté et développé ce département d'exception, à mes collègues thésards Maël Amari, Marie-Céline Bézat, Fabrice Clochard, Lucille Dompnier, et Sandrine Fisher pour le soutien mutuel dans les hauts et les bas

qui ont jalonnés nos thèses, à Virgile Charton, Stéphanie Donnet, Laurent Kirsche et Philippe Petit-Jean pour les échanges et collaborations de grande qualité, à Emmanuelle Diaz et Justin Parchantour pour l'accompagnement (notamment...) dans les traitements statistiques, à Emilie V., François C., JeanMi P., Laurence R., Olivier L., Sylvain D., Thomas B. ou encore Yann R. pour les bons moments passés dans l'open-space 516 mais aussi après le travail, ainsi qu'à la bande de « Vole Avec Tes Mains » (XiXi, Flo, Thibaut,...) pour les concerts sympathiques.

Je remercie les membres du Laboratoire des Images et Signaux (Gipsa-Lab, Institut National Polytechnique de Grenoble) pour leur accueil et leur disponibilité, et plus spécifiquement Jeanny Hérault et Nathalie Guyader pour leurs conseils lors des expérimentations et de mes visites mensuelles au laboratoire, ainsi qu'à Tien Ho Phuoc pour ses mesures et ses algorithmes.

Mes remerciements sont aussi adressés aux membres du Laboratoire Communication Langagière et Interaction Personne Système (Laboratoire d'Informatique de Grenoble) pour m'avoir accueilli lors de ma première année de thèse.

Ce travail n'aurait pas été ce qu'il est sans les riches échanges qui ont eut lieu notamment lors des conférences ECEM et ECVP. Je remercie en particulier Benjamin Tatler, Mark Wexler, Laurent Itti, Claus-Christian Carbon, Frédéric Landragin, Thérèse Collins, Alexandra Ciaccia, Tania Dukic, et Stefan Raidt.

Enfin, parce que sans eux rien ne serait pareil et tout serait moins bien, ces trois années de doctorat sont dédiées à ma famille et à mes proches, dont les qualités sont telles que les lister serait réducteur.

Avant Propos

Ce doctorat fait partie des 10% de thèses initiées chaque année en France sur des problématiques industrielles, et se déroulant au sein d'une entreprise via une Convention Industrielle de Formation par la Recherche (CIFRE). Le groupe PSA PEUGEOT CITROEN, par sa dimension internationale, par son activité fortement concurrentielle, et par le caractère technologique de ses produits, est doté d'une Direction Recherche et Innovation d'envergure. C'est ainsi que 20 nouvelles thèses sont financées chaque année, et qui par leur caractère amont sont de réels atouts pour l'innovation. La recherche en contexte industriel possède plusieurs caractéristiques qui différencient les thèses industrielles des thèses universitaires. Tout d'abord le positionnement et l'évolution de la problématique sont cadrés par la nécessité de transférer les travaux à des sujets concrets et porteurs, ce qui favorise les activités de synthèse pluridisciplinaires plutôt que les approfondissements théoriques. Dans certaines situations le doctorant peut être sollicité au sein de l'entreprise en tant qu'expert sur des projets connexe permettant de garder à l'esprit les contraintes industrielles, de suivre les orientations générales des innovations, et si nécessaire de repositionner le sujet sur un axe plus porteur. Une autre différence entre les thèses industrielles et universitaires provient du fait que les moyens expérimentaux alloués en milieu industriel sont généralement plus importants qu'en laboratoire universitaire. En contrepartie, le programme de recherche doctoral suit un plan rigoureux et balisé, dans lequel un esprit de synthèse et d'adaptation aux différents interlocuteurs est attendu. Une autre spécificité de la recherche industrielle vient du caractère confidentiel des sujets abordés, qui implique que les publications et communications vers la communauté scientifique soient contrôlées. Enfin, le doctorant « CIFRE » doit développer une expertise spécifique venant du co-encadrement universitaire et industriel : le premier favorise la compréhension des processus théoriques du sujet, et le second oriente le travail en faveur du transfert et de l'innovation en se basant sur des compétences pluridisciplinaires plutôt que spécialisées.

La synergie des différents acteurs et l'esprit d'ouverture de chacun auront ainsi permis le déroulement de ce travail dans d'excellentes conditions.

Sommaire

INTRODUCTION.....	1
VOLET 1 : MODELES ET PROCESSUS DE LA COGNITION VISUELLE	
CHAPITRE 1 PERCEPTION VISUELLE ET MOUVEMENTS OCULAIRES	5
1.1 PRINCIPES PHYSIOLOGIQUES	5
1.1.1 <i>Physiologie du système visuel.....</i>	<i>5</i>
1.1.2 <i>Classification des mouvements oculaires.....</i>	<i>8</i>
1.1.3 <i>Un monde décrit par 4 dimensions.....</i>	<i>11</i>
1.2 MOUVEMENTS OCULAIRES: TECHNIQUES ET APPLICATIONS	12
1.2.1 <i>Introduction.....</i>	<i>12</i>
1.2.2 <i>Techniques de mesure des mouvements oculaires.....</i>	<i>12</i>
1.2.3 <i>Paradigmes d'études des mouvements oculaires pour l'encodage des informations de scène.....</i>	<i>14</i>
1.2.4 <i>Paradigmes d'études des mouvements oculaires pour l'indexation des informations de scène.....</i>	<i>18</i>
1.2.5 <i>Synthèse.....</i>	<i>19</i>
CHAPITRE 2 SCENE, INFORMATION ET VISION.....	21
2.1 DE LA SCENE A SA REPRESENTATION DEPEINTE.....	21
2.1.1 <i>Définition du concept de « scène ».....</i>	<i>21</i>
2.1.2 <i>Différences entre scène réelle et scène dépeinte.....</i>	<i>22</i>
2.1.3 <i>Un comportement visuel propre à la nature réelle ou dépeinte de la scène.....</i>	<i>22</i>
2.2 PROPRIETES DES SCENES DEPEINTES	23
2.2.1 <i>Deux dimensions orthogonales : abstraction/concrétisation et stylisation/réalisme.....</i>	<i>23</i>
2.2.2 <i>Critères de distinction des scènes naturelles/artificielles.....</i>	<i>24</i>
2.2.3 <i>Définir la complexité des scènes visuelles.....</i>	<i>24</i>
2.2.4 <i>Synthèse.....</i>	<i>25</i>
2.3 DE LA SCENE AUX OBJETS QUI LA COMPOSE	25
2.3.1 <i>Structure holistique des informations de scène.....</i>	<i>25</i>
2.3.2 <i>Objets de scène, approche sémiotique.....</i>	<i>26</i>
2.3.3 <i>Objets de scène, approche cognitive.....</i>	<i>27</i>
2.3.4 <i>Objets de scène, approche enactive.....</i>	<i>28</i>
2.3.5 <i>Objets de scène, approche neuro-perceptive.....</i>	<i>28</i>
2.3.6 <i>Un cas particulier: dissociation entre la perception de l'objet et de la scène.....</i>	<i>29</i>
CHAPITRE 3 MECANISMES COGNITIFS DE L'EVALUATION VISUELLE	31
3.1 MEMOIRES VISUELLES ET PROCESSUS D'IDENTIFICATION DES SCENES.....	31
3.1.1 <i>Introduction.....</i>	<i>31</i>
3.1.2 <i>Processus de mémorisation des scènes et objets de scène.....</i>	<i>31</i>
3.1.3 <i>Processus d'identification des scènes.....</i>	<i>36</i>
3.2 ATTENTION VISUELLE ET COMPORTEMENT DU REGARD	42
3.2.1 <i>Propriétés de l'attention visuelle.....</i>	<i>42</i>
3.2.2 <i>L'attention visuelle peut être guidée par des facteurs endogènes et exogènes.....</i>	<i>45</i>
3.2.3 <i>Orientation de l'attention visuelle par des facteurs de saillance visuelle.....</i>	<i>47</i>
3.2.4 <i>Orientation de l'attention visuelle par des mécanismes cognitifs.....</i>	<i>50</i>
3.2.5 <i>Modèles attentionnels intégratifs.....</i>	<i>54</i>
3.3 DE L'IDENTIFICATION A L'EVALUATION.....	58
3.3.1 <i>L'évaluation par l'attitude.....</i>	<i>58</i>
3.3.2 <i>L'évaluation par les attributs.....</i>	<i>59</i>
3.3.3 <i>Effet du contexte sur l'évaluation d'un objet.....</i>	<i>62</i>
3.3.4 <i>Effets de l'implication et de l'expertise sur l'évaluation d'un objet.....</i>	<i>63</i>
3.3.5 <i>Influence de l'aisance du traitement de l'information sur l'évaluation.....</i>	<i>64</i>
3.3.6 <i>Préférence, appréciation, nouveauté, typicité : des interactions « complexes ».....</i>	<i>67</i>
3.3.7 <i>Conclusion.....</i>	<i>72</i>

VOLET 2 : PARTIE EXPERIMENTALE

CHAPITRE 4	CADRE EXPERIMENTAL	75
4.1	PRESENTATION	75
4.2	SPECIFICITES DES PARADIGMES EMPLOYES	77
CHAPITRE 5	TRAITEMENT DES DONNEES OCULOMETRIQUES	81
5.1	ACQUISITION DES DONNEES ET PRE TRAITEMENTS	81
5.2	ESTIMATION ET EXPLOITATION DES DISTRIBUTIONS DE FIXATIONS : DENSITES ET CLASSIFICATIONS	83
5.2.1	<i>Estimation des fonctions de densité : distribution spatiale des fixations du regard sur la scène</i>	<i>83</i>
5.2.2	<i>Méthodes de comparaison des fonctions de distributions des fixations.....</i>	<i>87</i>
5.2.3	<i>Méthode de classification et de regroupement des points de fixation</i>	<i>89</i>
5.3	ANALYSE DES ZONES D'INTERET	90
5.4	IDENTIFICATION DE STRATEGIES D'EXPLORATION VISUELLE.....	94
5.5	EVALUATION DES FACTEURS PREDICTIFS DU COMPORTEMENT OCULOMOTEUR	94
5.6	SYNTHESE	95
CHAPITRE 6	EXPERIMENTATION 1 « PARADIGME D'EVALUATION MONADIQUE »	97
6.1	MATERIEL ET METHODES.....	97
6.1.1	<i>Objectifs.....</i>	<i>97</i>
6.1.2	<i>Matériel</i>	<i>97</i>
6.2	RESULTATS	99
6.2.1	<i>Introduction : Jugement des produits</i>	<i>99</i>
6.2.2	<i>Analyse des données oculométriques : fonctions de densité spatiale des fixations.....</i>	<i>100</i>
6.2.3	<i>Analyse du comportement du regard par zones d'intérêt.....</i>	<i>111</i>
6.2.4	<i>Classification des stratégies oculomotrices.....</i>	<i>125</i>
6.2.5	<i>Conclusion.....</i>	<i>127</i>
CHAPITRE 7	EXPERIMENTATION 2 : PARADIGME D'EVALUATION PAR PAIRE	129
7.1	MATERIEL ET METHODES.....	129
7.1.1	<i>Objectifs.....</i>	<i>129</i>
7.1.2	<i>Matériel</i>	<i>129</i>
7.2	RESULTATS	132
7.2.1	<i>Jugement des produits</i>	<i>132</i>
7.2.2	<i>Analyse descriptive des fonctions de distribution spatiales des fixations oculaires</i>	<i>134</i>
7.2.3	<i>Analyse comparative des distributions de fixations</i>	<i>140</i>
7.2.4	<i>Analyse des processus temporels de l'attention visuelle sur la scène.....</i>	<i>142</i>
7.2.5	<i>Analyse du traitement visuel des zones d'intérêt.....</i>	<i>147</i>
7.2.6	<i>Existence de stratégies globales interindividuelles</i>	<i>154</i>
7.2.7	<i>Confrontation de la verbalisation à la prise d'information.....</i>	<i>155</i>
7.3	CONCLUSION.....	157
CHAPITRE 8	PARADIGME DE VISION CONTRAINTÉ	159
8.1	MATERIEL ET METHODES.....	159
8.1.1	<i>Présentation.....</i>	<i>159</i>
8.1.2	<i>Description du paradigme</i>	<i>159</i>
8.2	OBSERVATIONS.....	161
8.2.1	<i>Effets de la contrainte de la position du regard</i>	<i>161</i>
8.2.2	<i>Evaluation monadique</i>	<i>161</i>
8.2.3	<i>Présentation par paire : évaluation de la sportivité</i>	<i>162</i>
8.2.4	<i>Synthèse.....</i>	<i>163</i>
CHAPITRE 9	PREFERENCE ET ORIENTATION DU REGARD.....	165
9.1	PRESENTATION	165
9.2	RESULTATS	168

9.2.1	<i>Etude du temps nécessaire à la prise de décision</i>	168
9.2.2	<i>Manifestation oculomotrice de la prise de décision</i>	168
9.2.3	<i>Spécificités de la consigne « préférence</i> ».....	170
9.2.4	<i>Exploration visuelle de la figure choisie</i>	171
9.2.5	<i>Homogénéité oculomotrice inter-essais</i>	172
9.2.6	<i>Effet oculomoteur « idiosyncrasie sujet »</i>	174
9.2.7	<i>Etude complémentaire</i>	175
9.3	SYNTHESE.....	176
CHAPITRE 10 EVALUATION DES FACTEURS ATTENTIONNELS		178
10.1	PRESENTATION.....	178
10.2	PARADIGME DES « BUBBLES ».....	178
10.2.1	<i>Présentation</i>	178
10.2.2	<i>Résultats</i>	180
10.3	EXPLORATION LIBRE ET SAILLANCE VISUELLE.....	182
10.4	EVALUATION DES FACTEURS ATTENTIONNELS SUIVANT UN MODELE DE MELANGE ADDITIF.....	183
10.4.1	<i>Présentation</i>	183
10.4.2	<i>Résultats sur les distributions des fixations issues de l'exploration libre</i>	186
10.4.3	<i>Résultats sur les distributions des fixations issues des expérimentations avec consignes</i>	188
CHAPITRE 11 SYNTHÈSE DE LA PARTIE EXPERIMENTALE		194
DISCUSSION/PERSPECTIVES		198
REFERENCES		207

ANNEXES : DOCUMENTS COMPLEMENTAIRES CONFIDENTIELS

Introduction générale

Les facteurs humains sont de plus en plus profondément intégrés aux processus de conception industrielle, pour une meilleure prise en compte des interactions entre le produit et ses usagers. Les mécanismes perceptifs sont ainsi étudiés pour orienter favorablement les spécifications techniques et le fonctionnement cognitif humain est analysé pour optimiser la prise en main et l'usage des produits industriels. Depuis la fin des années 90, la société PSA Peugeot Citroën étudie la perception visuelle et le comportement oculomoteur pour la conception et la validation des systèmes d'informations embarqués (problématiques de vigilance, de navigation), des systèmes d'assistance à la conduite (détection des états de d'hypovigilance, lien entre la détection du danger et les contrôles visuels, etc.), et pour une meilleure compréhension des activités de conduite. Ce type d'exploitation des mesures du comportement du regard (oculométrie) effectuée par des ergonomes et des psychocogniticiens, associée aux techniques de recueils verbaux (notes, questionnaires, entretiens), consiste à étudier le lien entre la cognition visuelle et le comportement oculaire. Dans le même temps, alors que les capacités techniques des oculomètres (« eye-tracker », outil permettant de suivre les mouvements du regard) évoluaient très rapidement, de nombreux domaines applicatifs émergeaient sur la base de la mesure de la position du regard : navigation dans un système de réalité augmentée, environnements immersifs interagissant avec le regard, ou systèmes de vision artificielle bio-inspirée. Ainsi, les mesures du comportement du regard sont exploitées tant par les psycholinguistes qui observent la relation entre la position de l'œil en situation de lecture et le langage, par les ergonomes qui s'intéressent aux comportements visuels en situation de conduite ou lors de la recherche d'informations sur des sites web, qu'aux ingénieurs qui cherchent à optimiser en temps réel la qualité des images en fonction des zones visuelles qui attirent le regard.

Dans ces différentes situations, l'orientation du regard est fortement corrélée à l'orientation de l'attention visuelle, qui est elle-même guidée par les buts de tâche. Ainsi, presque toutes les zones fixées par le regard sont utilisées pour la résolution de la consigne. L'oculométrie permet alors de dégager le traitement cognitif séquentiel des informations visuelles traitées, étant ainsi, selon la formule consacrée, une fenêtre ouverte sur l'esprit.

L'analyse des interactions perceptives et cognitives entre un objet et son utilisateur est ainsi une phase incontournable dans le cycle de conception et de développement des produits manufacturés. Or, au-delà des spécifications liées à l'usage, la multiplicité de l'offre commerciale et la haute exigence des clients impliquent de définir avec soin les caractéristiques sensorielles et sémantiques d'un produit et d'identifier les critères d'évaluation afin de produire une satisfaction élevée et large. La vision est l'une des principales modalités sensorielles permettant d'acquérir des informations sur un objet, puis de l'évaluer. Ainsi, une première impression visuelle positive va le rendre attractif alors qu'une première impression négative va produire un rejet, une impression de complexité risque de repousser une partie des clients et un caractère très innovant risque de n'être apprécié que par une partie la cible. Il est donc intéressant d'identifier les caractéristiques visuelles sur lesquelles s'élabore le jugement afin d'orienter favorablement l'évaluation du produit, en fonction des spécificités cognitives des clients (expertise, attente, etc.). Cependant, les recueils classiques tels que la verbalisation *a posteriori* ou en phase d'observation ne semblent pas refléter réellement les informations visuelles traitées car de nombreux filtres interviennent : encodage sélectif des informations visuelles, barrière du langage, filtre de la conscience. Il a alors été envisagé de transférer les techniques oculométriques à l'étude des processus d'évaluation visuelle. Parce que le regard semble synchronisé à l'activité cognitive et que la qualité des informations visuelles est meilleure au centre du champ visuel qu'en

périphérie, on peut en déduire que l'observation du comportement oculomoteur permet de mesurer « à la source » les informations visuelles traitées pour l'évaluation. La dynamique et la position du regard va alors servir à inférer les processus perceptifs et cognitifs activés par l'évaluation.

Cependant, le comportement oculomoteur se révèle être soumis à de grandes variabilités lorsque l'on ne se trouve pas dans des situations où l'attention est fortement orientée par la tâche, c'est-à-dire en dehors des activités utilisant classiquement l'oculométrie. En effet, les processus attentionnels n'apparaissent pas être de même nature en situation de conduite que lors de l'évaluation visuelle d'un objet. Dans ce dernier cas, l'attention visuelle semble traiter les informations visuelles en parallèle sous l'influence de multiples processus, et le regard pourrait être guidé tant par des facteurs attentionnels exogènes (orientation automatique du regard par les propriétés de la scène visuelle : zones de contours, zones de discontinuités, de forte brillance), que par des facteurs endogènes (expression symbolique, incohérences sémantiques, éléments d'identification de la scène, etc.). Ainsi, le comportement du regard n'est plus directement lié à l'attention, qui est elle-même influencée par d'autres processus que les buts de tâche. Cela explique alors la très grande variabilité des comportements visuels. Il est donc nécessaire de mettre au point des méthodes d'analyse de données statistiques pour dégager les effets des différents attracteurs du regard, pour lier prise d'information et jugement et pour dégager le cas échéant des groupes de stratégies tant de prise d'information que de résolution de la tâche.

Puisque les comportements oculomoteurs s'avèrent peu consistants entre eux et indirectement connectés aux conditions expérimentales, PSA PEUGEOT CITROEN, dans sa démarche d'étude des processus évaluatifs, a soulevé les problématiques suivantes : quels sont les processus attentionnels activés lors d'une tâche d'évaluation, comment traiter les données oculométriques pour mettre en évidence ces processus, et quel potentiel contient la mesure oculométrique pour l'aide à la compréhension des mécanismes de jugement. Ces problématiques étant amont à l'exploitation des mesures oculométriques dans le contexte d'évaluation des produits, ce présent projet de thèse a été monté sous forme d'une convention industrielle de formation par la recherche (CIFRE) entre PSA PEUGEOT CITROEN, sous le tutorat de Pauline Faye (département « PErception et Facteurs Humains »), l'Agence Nationale pour la Recherche Technique via l'Ecole Doctorale pour la Santé, la Cognition et l'Environnement de Grenoble, et sous la supervision de Anne Guérin-Dugué (Gipsa-Lab /ex. L.I.S./ de l'Institut National Polytechnique de Grenoble), de Michel Dubois (Laboratoire de Psychologie Sociale, Grenoble) et de Christian Marendaz (Laboratoire de Psychologie et NeuroCognition, Grenoble).

Cette thèse, qui s'est déroulée de janvier 2005 à décembre 2007 dans les locaux de la société PSA PEUGEOT CITROEN s'est naturellement positionnée dans le domaine émergent et pluridisciplinaire des sciences cognitives. En effet, les recherches en perception visuelle font appel à de nombreuses disciplines tant des sciences de l'ingénieur que des sciences humaines: les processus de traitement des informations visuelles sont étudiés en collaboration entre les neuropsychologues et les ingénieurs du traitement du signal, les processus de reconnaissance des scènes sont étudiés par les psychologues cogniticiens et par les chercheurs en sciences de l'information, les psychosociologues étudient les comportements du regard pour la reconnaissance des émotions avec des ingénieurs établissant des systèmes automatiques de reconnaissance des expressions faciales. Dans le champ des sciences cognitives, notre problématique se positionne dans la discipline de l'analyse statistique des données et des signaux provenant de mesures comportementales, appliquée à l'identification des processus d'attention visuelle et de formation d'impression. Ce travail de thèse se veut donc un apport méthodologique d'exploitation de la mesure des mouvements du regard lors de l'évaluation visuelle de produits manufacturés, par la conception et l'application de méthodes d'analyse des données oculométriques, par la construction de paradigmes expérimentaux permettant de dégager les

processus d'attention, et par une observation expérimentale des liens entre les processus de prise d'information visuelle et le jugement du produit automobile.

La première partie de ce manuscrit est une étude théorique et bibliographique des processus impliqués dans la prise d'information visuelle et la formation d'impression. Dans un premier chapitre, nous effectuons un bilan des principes physiologique et biologique du système visuel, afin de dégager les fonctions de bases du comportement d'orientation du regard. Puis nous nous intéressons aux moyens de mesure du regard et aux différentes exploitations de cette mesure, tant pour des recherches théoriques qu'applicatives. Le second chapitre présente les différents types de scènes visuelles au regard des spécificités des stimuli que nous allons étudier. Les niveaux d'informations contenues dans une scène visuelle sont définis. Un tour d'horizon des mécanismes cognitifs impliqués dans le traitement visuel des scènes est effectué dans le troisième chapitre. La première section porte sur les différents types de mémoire visuelle et sur les processus d'identification des scènes. La seconde section aborde les différents types d'attentions de la cognition visuelle. Les liens entre l'orientation du regard et l'attention y sont analysés. Enfin la troisième section présente les principaux processus qui sous-tendent le jugement et la formation d'impression, ainsi que les biais d'évaluation dus à l'aisance au traitement des informations visuelles.

La seconde partie de ce document présente les travaux expérimentaux qui ont été réalisés dans cette thèse. Le premier chapitre présente les paradigmes qui sont développés et les spécificités de nos facteurs expérimentaux. Le second chapitre présente les méthodes de traitement des données oculométriques qui sont appliquées pour l'analyse des résultats expérimentaux. Le troisième chapitre présente la méthode et les résultats d'une expérimentation d'évaluation monadique d'habitacles de véhicules. Il est directement suivi par un chapitre exposant la méthode et les résultats d'une expérimentation d'évaluation par paire. Une expérimentation complémentaire utilisant l'oculomètre non plus en tant que système de mesure mais système de contrainte est proposée dans le cinquième chapitre. Le sixième chapitre aborde le rôle de pointeur affectif et cognitif que manifeste le regard en situation d'évaluation d'objets, et vérifie l'existence de biais dépendants de l'orientation du regard dans l'évaluation visuelle d'objets. Le septième chapitre est une synthèse des différents résultats expérimentaux permettant de mieux comprendre les processus attentionnels et comportementaux observés dans nos expériences. La première section présente la méthode dite des « Bubbles » qui apparaît complémentaire aux mesures oculométriques pour relever les zones visuelles de décision; la seconde section expose une étude sur le comportement du regard en situation d'exploration libre du produit automobile ; enfin la troisième section propose une méthode d'évaluation de la contribution de différents facteurs attentionnels sur la distribution des fixations oculaires. Le dernier chapitre synthétise les différents résultats expérimentaux et les méthodes appliquées, pour conclure et proposer les perspectives consécutives à ces travaux.

- Volet 1 -
MODELES ET PROCESSUS DE LA
COGNITION VISUELLE

Chapitre 1

Perception visuelle et mouvements oculaires

1.1 Principes physiologiques

L'étude des processus de la perception visuelle se doit d'être accompagnée d'une compréhension des principes anatomiques et biologiques du système visuel. En effet, ce système visuel est composé de capteurs ayant des propriétés spécifiques et d'unités de traitement permettant l'optimisation de l'extraction des informations visuelles; c'est ainsi que la densité des capteurs sur la rétine explique la non homogénéité de résolution du champ visuel et par là même la nécessité des mouvements des yeux. Dans cette partie nous rappellerons tout d'abord les propriétés physiologiques de l'œil, puis nous aborderons les différents mouvements oculaires répertoriés dans la littérature, et enfin nous nous intéresserons à la manière dont les différentes informations sont assemblées pour permettre la perception de la profondeur.

1.1.1 Physiologie du système visuel

L'œil humain est un système complexe possédant de nombreuses fonctions optiques, permettant l'acquisition de flux d'informations visuelles (Figure 1).

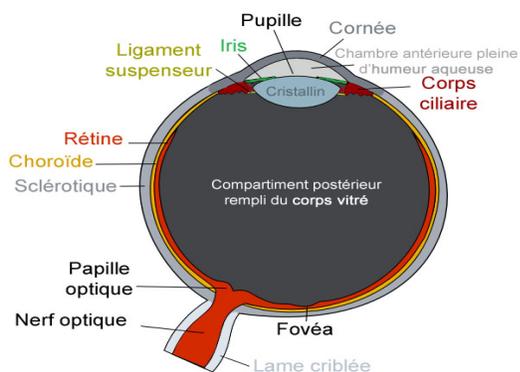


Figure 1 Structure physiologique de l'œil¹

Outre les muscles qui permettent d'orienter et de focaliser le regard (mouvements horizontaux, verticaux, et vergence), l'œil possède un système de récepteurs photosensibles, implantés sur la rétine. Trois principales couches composent la rétine (Figure 2) : l'épithélium pigmentaire, dont la

¹ Source : <http://www.wikipedia.org>

fonction est de fournir le rétine à aux photorécepteurs, la rétine sensorielle, ayant un rôle de transduction des influx lumineux et composée des photorécepteurs, et la rétine cérébrale, traitant les influx nerveux.

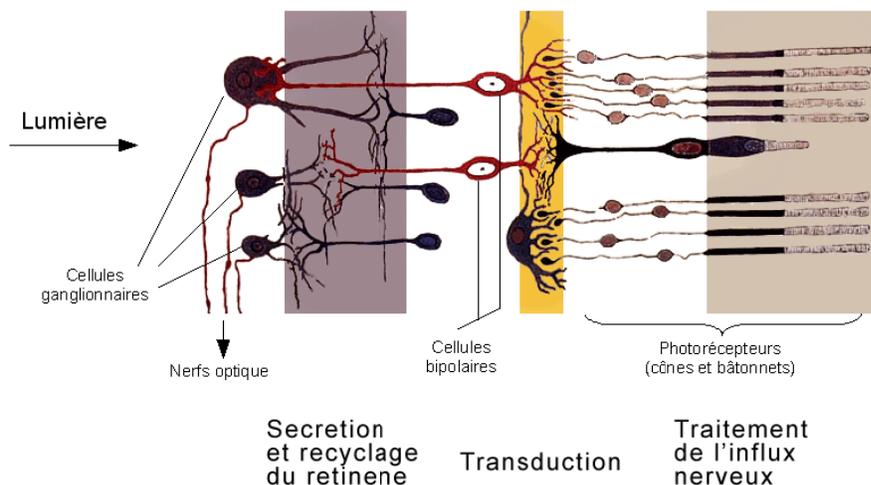


Figure 2 Rétine humaine: structure laminaire de la rétine²

Il existe deux principaux types de photorécepteurs: les cônes et les bâtonnets, dont les répartitions sur la rétine sont inégales. Les bâtonnets, au nombre de 100 à 200 millions, sont absents de la zone centrale du champ visuel nommé « fovéa », qui est la partie de la rétine où la résolution de la vision est la meilleure. Le rôle des bâtonnets est de fournir des informations de luminance, permettant une vision nocturne. Ces informations sont à faible résolution spatiale adaptées à la perception des mouvements. Les cônes, au nombre de 5 millions, sont disposés au centre de la rétine (fovéa) et couvrent un champ visuel d'environ $2 \times 2^\circ$. Ils ont pour fonction de fournir des informations visuelles à haute résolution spatiale et à sensibilité de chrominance élevée. Trois types de cônes sont identifiés : les cônes L (large, $\lambda=565\text{nm}$), M (middle, 535nm) et S (short - 430nm), selon la longueur d'onde lumineuse à laquelle ils sont sensibles (fig. 1-2).

² Source : <http://www.wikipedia.org>, tiré de [Cajal11]. Voir aussi [Kolb01].

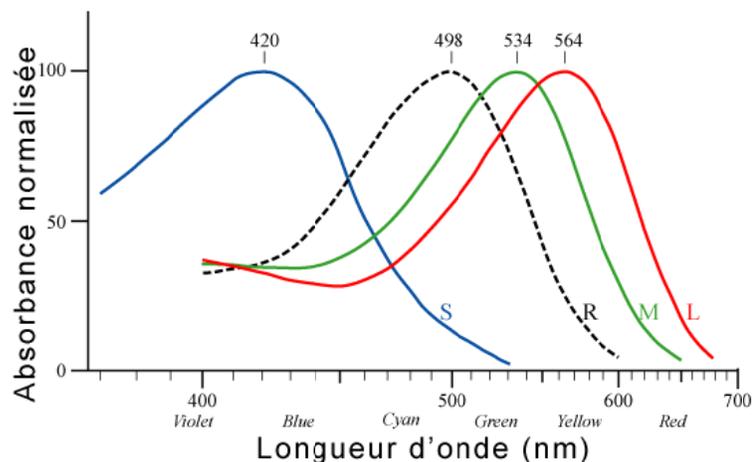


Figure 3 Courbe d'absorbance en fonction de la longueur d'onde. (S), (M), et (L) correspondent respectivement aux courbes pour les cônes bleus, verts et rouges, et (R) correspond à la courbe relative aux bâtonnets [Bowmaker80]

Ces différents cônes n'étant pas répartis équitablement sur la surface, la sensibilité visuelle n'est pas égale sur l'ensemble du champ de vision pour toutes les longueurs d'onde (Figure 3). Le rôle des cônes est maximisé dans la vision de jour. Notons que la sensibilité aux contrastes est elle aussi inégale sur l'ensemble du champ visuel, car étant faible pour les récepteurs périphériques ([Koenderink99], [Kelly84]).

Tel que nous l'avons vu dans le paragraphe précédent, la fovéa est la partie centrale circulaire de la rétine couvrant environ 4° , et dont le centre (la macula) couvre 2° (Figure 4).

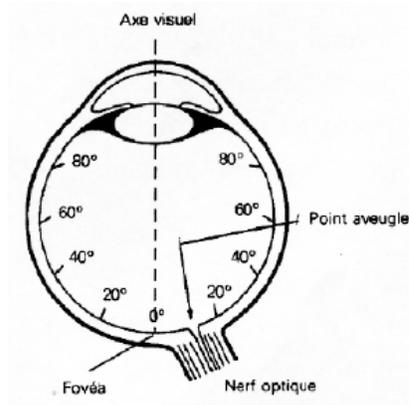


Figure 4 Schéma de l'axe optique de l'œil [Lariani00]

La surface de cette zone du champ visuel correspond environ à la largeur d'un pouce tendu à bout de bras. La para-fovéa couvre 15 à 20° du champ visuel. Il s'agit de la zone de recouvrement entre les bâtonnets et les cônes. La périphérie est la zone complétant le champ visuel permettant de couvrir 90° horizontaux et 70° verticaux. La résolution visuelle de cette zone est de moins de 50% de celle de la fovéa (Figure 5) [Jacob95].

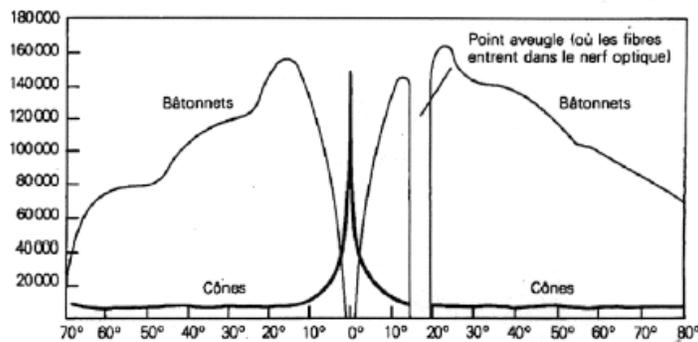


Figure 5 Distribution de la densité des cônes et bâtonnets sur la rétine [Lariani00]

Le champ visuel central est défini comme l'espace à résolution maximale que perçoit l'œil quand le regard est fixe.

✓ Synthèse

La résolution et la sensibilité en chrominance et luminance de la vision étant spatialement non homogènes, on comprend la nécessité d'un mécanisme complémentaire pour l'acquisition des informations visuelles sur le monde qui nous entoure [Bruce90]. Le système visuel doit s'adapter afin de traiter avec un niveau de résolution adéquat chaque information du champ visuel. Le rôle de base des mouvements oculaires est ainsi de déplacer la zone du champ visuel où la résolution est la meilleure en direction des informations visuelles à traiter, et cela sans avoir à tourner la tête. Le chapitre suivant présente ainsi les différents mouvements de l'œil.

1.1.2 Classification des mouvements oculaires

1.1.2.1 Pourquoi bouger les yeux

Les mouvements oculaires peuvent se justifier par au moins trois facteurs. Le premier, présenté précédemment, consiste en une distribution non homogène des capteurs rétinien, impliquant la nécessité de repositionner le champ visuel pour un traitement optimal des informations visuelles. Le deuxième facteur renvoie à la capacité de traitement par le système cognitif des informations visuelles, qui pourrait ne pas être assez puissant pour traiter simultanément la totalité des informations présentes dans le champ visuel. La dégradation spatiale de la résolution visuelle permet ainsi de filtrer physiologiquement une partie des informations. Les mouvements oculaires trouveraient leur raison d'être en tant que filtres de l'information visuelle sur demande du système cognitif. Le troisième facteur justifiant les mouvements des yeux provient du fonctionnement par potentiel d'action du système nerveux. L'existence permanente de micro saccades rafraichirait l'influx nerveux stimulant les neurones de la vision, afin de permettre une perception continue.

1.1.2.2 Comment bouger les yeux

Les mouvements des yeux s'expliquent par la disposition des muscles oculaires.

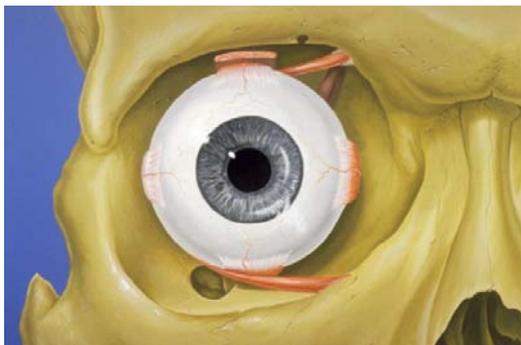


Figure 6 Anatomie normale de l'œil humain et de l'orbite, vue antérieure [Lynch06]

L'œil est contrôlé par trois ensembles musculaires : le premier pour les mouvements de rotation (« rolling »), le deuxième pour les mouvements verticaux, le troisième pour les mouvements horizontaux. La littérature ([Jacob95], [Bruce90]) retient sept types de mouvements oculaires.

✓ *Vergence (ajustement à la profondeur de champ)*

La vergence est un mouvement synchronisé des deux yeux, dont les directions respectives convergent. Cela permet de fixer un objet par les deux yeux, même lorsque la distance entre l'observateur et l'objet évolue. Un contrôle volontaire de ces muscles permet la vision des stéréogrammes³ [Humphrey89]. La vitesse est au maximum de 10°/sec.

✓ *Rotation*

Comme son nom l'indique, la rotation est un mouvement de l'œil tournant sur lui-même.

✓ *Nystagmus physiologique*

Ce mouvement à haute fréquence oscillatoire produit un saut de l'image sur la rétine, permettant de rafraîchir les informations traitées par les récepteurs rétinien. Ainsi, lors de l'absence pathologique de ce mouvement, l'image visuelle disparaît, rendant le sujet aveugle en quelques secondes jusqu'au prochain mouvement de tête. Le déplacement doit être équivalent à la distance séparant deux cônes fovéaux dans une durée de 100ms maximum. Il s'agit d'un mouvement involontaire de moins de 1°.

✓ *Fixations et Saccade : scanner le champ visuel*

Il s'agit du principal mouvement permettant de recentrer rapidement le regard sur différentes zones du champ visuel. Ce mouvement consiste en un déplacement soudain et rapide des yeux. En moyenne, un individu effectue plus de 150 000 mouvements saccadiques des yeux par jour [Abrams92]. Le temps de latence (de l'apparition du stimulus au mouvement) est d'au minimum 100ms, et la durée de la saccade varie entre 30 et 120ms. Les saccades peuvent être volontaires ou involontaires, et sont de type balistique, c'est-à-dire que lorsque le mouvement est initié, il n'est plus possible de le stopper, de modifier son trajet ou sa destination. Le traitement de l'information visuelle est inhibé pendant le mouvement [Reingold02] (et les 50ms qui le précèdent) ce qui explique que l'on ne perçoive ni flou ni illusion de déplacement de la scène lorsque la position du regard change brusquement. Les périodes qui séparent deux saccades sont appelées des fixations, et durent généralement entre 200 et 600ms. La vitesse de déplacement de l'œil peut atteindre 600°/sec. Ainsi

³ Stéréogramme : représentation picturale spécifique construite à partir de deux points de vue d'une même scène, et permettant la vision en relief de cette scène

un angle de 5° est parcouru en environ 30ms et un angle de 40° en 100ms. La combinaison des mouvements des yeux sous forme de fixations et de saccades, associée aux mouvements de tête, produit une illusion de vision à haute résolution continue dans l'espace et le temps. Notons d'ailleurs qu'il a été récemment observé que la perception du temps est ralentie lors des mouvements saccadiques du regard [Morrone05]. Dans notre travail nous nous intéresserons uniquement à ce type de mouvements oculaires.

✓ *Sauts et micro saccades*

Ils apparaissent pendant les fixations, et consistent en des sauts lents suivis de toutes petites saccades. Leur utilité n'a pas été établie. Il s'agit de mouvements involontaires qui pourraient provenir d'instabilités des flux nerveux.

✓ *Mouvement de poursuite (suivi d'objets mobiles)*

Il s'agit d'un mouvement plus lent et lisse que la saccade. Il est utilisé pour fixer objet qui se déplace. Il semble difficile de l'effectuer de manière volontaire. La vitesse maximale de l'objet est de 30°/sec.

✓ *Mouvement opto-cinétique*

Ce mouvement apparaît lors d'un mouvement de tête ou d'un mouvement répétitif (phénomène d'observation par la fenêtre d'un train). Il consiste en une poursuite lente dans une direction pour suivre un élément de la scène, suivi par un mouvement rapide dans la direction opposée pour sélectionner une nouvelle position.

1.1.2.3 Synthèse

Mouvement	Volontaire/involontaire	latence	Spécification
vergence	volontaire	lent	10°/sec
rotation	involontaire	rapide	
poursuite	involontaire	rapide	30°/sec
Opto cinétique	volontaire	lent	
nystagmus	involontaire	rapide	
Sauts et micro saccades	involontaire	lent	
saccades	volontaire	rapide	600°/sec

Tableau 1 Synthèse des différents mouvements oculaires

Les processus qui sous-tendent les mouvements oculaires sont particulièrement complexes [Wandell95] car de nombreux facteurs interviennent. La fonction de base du mouvement du regard est de rafraîchir l'image rétinienne en terme d'activité neuronale afin d'éviter la disparition des informations lumineuses en 1 à 3 secondes. Une dynamique du flux visuel est ainsi nécessaire pour une perception optimale. Une seconde fonction du mouvement est d'améliorer la perception des contrastes et des détails des informations du champ visuel en recentrant la fovéa sur les zones traitées. Le mouvement balistique saccadique permet donc d'extraire des informations spatiales à haute résolution sur la zone fovéale et de repositionner très rapidement le regard sur les zones d'intérêt de la scène. Du fait du faible coût énergétique requis et de la nature automatique du mouvement, nous ne percevons pas que notre acuité spatiale n'est pas uniforme sur l'ensemble du champ visuel. Ainsi, dans la vie quotidienne le regard est souvent orienté en direction des unités informationnelles pertinentes pour la tâche en cours, même si la haute résolution des informations visuelles n'est pas exploitée. Les mouvements involontaires des yeux (c'est-à-dire naturels et automatiques) pourraient ainsi être des indicateurs de l'activité cognitive [Liversedge00]. La section

suivante présente l'intérêt de la vision binoculaire pour une adaptation optimale au monde qui nous entoure, notamment pour l'acquisition des informations de profondeur et de mouvements.

1.1.3 Un monde décrit par 4 dimensions

Notre vie quotidienne se déroule dans un espace comportant des objets mobiles (mouvement : dimension temporelle) évoluant selon trois dimensions spatiales (largeur, longueur, profondeur). Cette affirmation, triviale, n'est cependant pas dénuée d'intérêt lorsque l'on souhaite étudier la perception visuelle. En effet, nos processus de traitement de l'information fournissent en un minimum de temps des informations pour une adaptation permanente à l'environnement. Ainsi, le mouvement d'un objet dans notre champ de vision périphérique sera automatiquement détecté et l'attention s'y portera dans un délai très bref, par exemple pour produire une réaction s'il existe un danger [Yantis 02]. Si les mouvements sont aisément détectés dans un but de vigilance, la mémoire aussi est adaptée à un traitement optimal d'un monde en mouvement. Si une scène réelle est observée, la forme des objets sera encodée de manière stable mais leur position de manière transitoire, ce qui ne se vérifiera pas sur l'observation de la photographie de cette même scène [Tatler05]. Le regard possède aussi une fonction d'anticipation de la programmation du mouvement : le regard accompagne, voir précède, l'action principale dans la vie quotidienne, observée par Hayhoe [Hayhoe03] sur des participants préparant un sandwich dans une cuisine. Le regard possède donc des fonctions de vigilance et des fonctions de coordination motrice, c'est à dire de pointage spatial en fonction du temps.

Un système mono-capteur ne peut mesurer la profondeur de champ d'une scène : la distance entre deux objets disposés dans le même axe de fuite n'est pas calculable. Ainsi, sans connaissance préalable du monde qui nous entoure et de ses échelles il est impossible de calculer l'éloignement d'un objet à partir de sa taille perceptive. Un objet se déplaçant vers l'observateur à vitesse constante dans le plan sagittal est alors analysé par le système visuel comme un point grossissant, sans autre information de distance que la taille relative du point.

La pré-connaissance du monde qui nous entoure est ainsi l'une des fonctionnalités permettant de traiter la troisième dimension. Le système visuel possède plusieurs manières d'extraire les informations de profondeur. On peut ainsi distinguer les voies proprioceptives, qui permettent d'appréhender la profondeur de champ à partir des propriétés physiologiques du système visuel, et les voies computationnelles, sur la base de reconstitutions géométriques.

Les voies proprioceptives se décomposent en au moins deux processus : l'accommodation et la vergence. Il s'agit de fonctions de mise au point permettant de rendre net les contours d'un objet quelle que soit sa distance (dans les limites physiologiques du *punctum remotum*⁴ et *proximum* cependant). Le traitement de la profondeur est alors possible jusqu'à ce que la taille des objets soit inférieure à la taille des photorécepteurs (classiquement 1° d'arc). L'accommodation permet d'extraire une information de profondeur par le relevé de la courbure du muscle agissant sur la netteté de l'image formée sur la rétine. La seconde fonctionnalité proprioceptive est la vergence, et s'applique à un système binoculaire. Chaque œil converge en direction de l'objet observé. L'angle entre les directions de chaque œil indique si l'objet est éloigné (faible déviation) ou proche (forte déviation).

La principale voie computationnelle de traitement de la profondeur est la stéréoscopie [Julesz86]. Chaque œil fournit au cerveau une image bidimensionnelle de la scène. Du fait de l'écartement des yeux, ces points de vue sont décalés de 5 à 8 cm. Cependant, plutôt que de percevoir deux images de

⁴ http://fr.wikipedia.org/wiki/Punctum_remotum : On appelle *punctum remotum* d'un œil, le point le plus éloigné que l'on peut observer distinctement. On appelle *punctum proximum* d'un œil, le point le plus proche que l'on peut observer distinctement.

la même scène, le cerveau reconstitue une représentation fusionnée de la scène où le décalage des positions des objets entre les deux images va être un indicateur du relief, du volume, et de la profondeur. Une simple expérimentation permet de vérifier cela. En pointant son doigt dans une direction précise de l'espace, par exemple vers un coin de la pièce, puis en fermant alternativement un œil puis l'autre, on perçoit que la distance entre le doigt et le repère n'est pas constante, et est encore différente lorsque les deux yeux sont ouverts. L'image reconstituée est donc bien une adaptation des deux images monoculaires. Notons que le décalage entre le repère et le doigt n'est pas le même selon l'œil ouvert, indiquant un processus de rivalité binoculaire. Cela signifie que le flux visuel d'un œil est privilégié par rapport à l'autre. Cet œil dont le flux visuel est privilégié est nommé œil dominant. C'est ce principe de vision binoculaire qu'exploitent les stéréogrammes.

Les processus qui permettent la vision stéréoscopique sont ainsi efficaces mais très complexes. Autant la mesure de convergence dans un plan permet d'extraire une information spatiale précisément localisée, autant la vision dans l'espace tridimensionnel est complexe à traiter, et encore plus à mesurer. Notre travail s'inscrit ainsi dans un espace bidimensionnel, c'est-à-dire sans prendre en compte les processus perceptifs liés à la profondeur de champ. Afin de mieux comprendre l'intérêt de la mesure des mouvements des yeux, le chapitre suivant présente d'une part les différentes méthodes permettant de mesurer le comportement du regard, mais aussi les différents usages utilisant ces mesures.

1.2 Mouvements oculaires: techniques et applications

1.2.1 Introduction

Cette partie présente différentes situations dans lesquelles la mesure du mouvement du regard est exploitée. De nombreuses disciplines utilisent désormais les systèmes oculométriques, avec cependant des approches théoriques très différentes. L'inférence des processus cognitifs par les mouvements des yeux est dépendante des hypothèses de traitement de l'information visuelle, propre à chaque discipline. La synthèse des usages abordée dans cette partie n'a pas pour but de présenter les modèles théoriques associés à chaque application, mais plutôt de spécifier le type d'informations recherchées dans les différentes conditions expérimentales. Les paradigmes se classent en deux catégories. La première exploite les mouvements des yeux à visée d'encodage, afin d'identifier les zones visuelles sur lesquelles s'est porté l'attention, parfois nommées paradigmes à usage diagnostique » [Duchowski03], et la deuxième regroupe les applications interactives, utilisant les mouvements du regard à usage d'index, nommé « paradigmes à usage interactif ».

1.2.2 Techniques de mesure des mouvements oculaires

L'usage de la mesure des mouvements des yeux dans un grand nombre de disciplines a produit le développement de différents systèmes de mesure. La plus ancienne et plus simple technique est l'observation directe du regard du participant. Cependant la vitesse de déplacement est telle qu'il n'y a que les lieux de départ et d'arrivée des saccades qui peuvent être observés. Wade & Tatler [Wade05] précisent que la résolution temporelle de la vision est faible, impliquant que même si l'expérimentateur observe attentivement, il lui est difficile de déterminer combien de temps les yeux du participant ont bougé. Dodge & Herdmann (1898, voir [Dodge1900]) ont ensuite mis au point un système d'observation par miroir du regard en situation de lecture. Cependant, seuls les mouvements les plus grands peuvent être observés. A partir de 1901 Dodge et Cline [Dodge1901] développent un système (« Dodge photochronograph ») qui consiste à enregistrer sur un film photosensible le reflet

d'une fente lumineuse sur la cornée. Un flash est activé toutes les 100ms, permettant ainsi d'imprimer sur une plaque photosensible 10 positions du regard par seconde. C'est avec un système basé sur ce principe que Buswell [Buswell35] effectuera ses travaux sur les patterns oculomoteurs en situation d'observation de peintures autour de 1935. En 1973, Cornsweet & Crane [Cornsweet73] développent un système qui utilise le mouvement relatif de la surface externe de la cornée avec celui de la surface interne de la lentille (Purkinje images – réflexion de lumière sur les surfaces internes et externes de la cornée et de la lentille), fournissant une mesure très précise du mouvement de l'œil.

L'arrivée de systèmes de mesure par lentilles de contacts via un champ électrique de surface va améliorer la précision des mesures du mouvement des yeux. Chaque mouvement, le plus petit soit-il, fait varier le potentiel de surface. Il est ainsi possible de mesurer avec précision le moindre mouvement oculaire. La principale qualité de ce type de système est la précision mais les grands défauts sont d'être invasifs et peu écologiques [McConkie02]; [Findlay97]. Une technique moins invasive et moins précise, mais cependant adaptée à la mesure des mouvements rapides est l'électro-oculographie. Cette technique consiste à mesurer, par le biais d'électrodes positionnées sur la peau aux alentours des yeux la différence de potentiel des influx nerveux aux alentours des muscles oculaires [Ilg93]; [Rosander04].

L'évolution conjointe des systèmes informatiques de forte puissance à des prix abordables avec les avancées des algorithmes et techniques de traitement d'image depuis les années 90 a permis des évolutions majeures des systèmes de mesure du regard. Tandis que les systèmes historiques contraignaient les mouvements de tête et nécessitaient l'usage de sources lumineuses gênantes pour les tâches de vision, les nouveaux systèmes deviennent non invasifs et portables. Deux types d'oculomètres existent : les systèmes distants et les systèmes à casque. Les premiers sont déportés à distance du participant (« remote eye-tracker ») et ont pour seule contrainte d'être positionnés dans le centre du plan frontal. Le principe de mesure n'est pas spécifiquement innovant car il s'agit toujours de recueillir la réflexion de cornée, mais la méthode employée est l'extraction algorithmique par reconnaissance de forme de reflets dans l'image. La mesure est optimisée par l'utilisation de sources infrarouges et est rendue plus robuste par la détection du contour de la pupille. Des caméras sensibles à l'infra rouge (longueurs d'ondes auxquelles notre système visuel est insensible) mesurent ces reflets de cornée et contours de pupille. Des algorithmes de traitement d'image extraient de l'image vidéo l'orientation des yeux. Leur convergence fournit la position du regard. L'avantage de cette méthode est la robustesse de la mesure (aux mouvements de regard et de tête) allié à une bonne précision. Son inconvénient principal est d'être sensible à la forme des yeux et du visage, ainsi qu'aux mouvements de tête, produisant des mesures de qualité variable. Ces systèmes sont, de plus, très sensibles aux sources infrarouges naturelles telles que l'éclairement du soleil, perturbant le traitement d'image.

Les oculomètres à casques ou oculomètres portables (« head mounted eye tracker ») sont constitués d'un casque sur lequel sont placés des petits miroirs en dessous des yeux, et de caméras mesurant le reflet émis par ces miroirs. Le flux vidéo est enregistré par un magnétoscope, les trames paires enregistrant les images de la caméra 1 et les trames impaires les images de la caméra 2. Il s'agit là aussi d'extraire la position du regard par reflet de cornée et contour de pupille. L'usage d'émetteur infra rouge de scène optimise la mesure [Land94]. L'avantage de ces systèmes (de nombreuses déclinaisons commerciales se partagent le marché) est de s'intégrer dans une activité écologique et d'être très précis lorsqu'ils sont correctement paramétrés. Leur principal inconvénient est d'être sensible aux variations lumineuses de l'environnement ainsi qu'aux vibrations et tressautements. Enfin, notons que ces systèmes à casque ne permettent pas en général d'obtenir une mesure absolue de la position du regard dans l'espace, mais la position du regard dans la scène courante. Cela implique que l'analyse du traitement des zones d'intérêt ne peut être que manuelle. Des systèmes récents permettent cependant de faire correspondre la position relative du regard dans

l'image et la position absolue par synchronisation entre l'oculomètre et un système de suivi spatial de la position et d'orientation de la tête (laser ou magnétique).

1.2.3 Paradigmes d'études des mouvements oculaires pour l'encodage des informations de scène

1.2.3.1 Le regard est synchronisé à l'activité mnésique

Le mouvement des yeux est parfois considéré comme étant un témoin de l'activité mentale et mnésique du participant. Le regard n'est pas ici un capteur sensoriel, mais un pointeur comportemental de l'activité cognitive. Une démonstration simple est présentée par Stern [Stern93] par le phénomène « d'orientations de la pensée visuelle » (orientation of thought looking). Ce processus apparaît lorsqu'il est demandé à un participant d'effectuer une tâche purement cognitive, sans qu'il ne porte son attention sur la scène visuelle. Il leur est demandé d'épeler le mot « MOTHER » à l'envers. La mesure du mouvement des yeux montre que le regard se déplace de droite à gauche au rythme de l'épellation comme si les participants visualisaient le mot et qu'ils lisaient simplement les lettres une à une. Le regard serait donc asservi à l'activité cognitive en tant que pointeur comportemental. Mais certains chercheurs vont plus loin, en proposant que le regard soit directement associé à la mémoire, en tant que capteur puis pointeur. C'est ainsi que Noton et Stark [Noton71a] ; [Noton71b] comparent les trajets du regard de participant entre deux phases : mémorisation d'une image puis reconnaissance de cette même image. Ils observent que les trajets seraient similaires en terme topologiques pour un même participant entre les deux phases. Ils suggèrent alors que la représentation interne de la scène visuelle est la conséquence de l'encodage d'une séquence mixte de traces de mémoires sensorielles et motrices, enregistrant alternativement les caractéristiques visuelles locales de la scène puis les caractéristiques du mouvement des yeux requis pour atteindre la caractéristique visuelle locale suivante.

Hebb [Hebb68] propose que les mouvements des yeux aient un rôle primordial dans la représentation mentale des scènes. Les mouvements oculaires seraient nécessaires à la collecte et l'organisation des portions d'images stockées en mémoire, afin de reconstituer une visualisation globale de la scène. Brandt et Stark [Brandt97] exposent à neuf participants des diagrammes aux motifs irréguliers et montrent que les mouvements oculaires spontanés qui se produisent pendant une phase de remémoration reflètent de manière précise le contenu et les relations spatiales de l'image originale. La remémoration de la scène produirait des mouvements des yeux suivant une topologie identique à ceux issus de l'encodage perceptif. Il apparaît cependant que la remémoration produit des fixations plus longues et des amplitudes de saccades plus faibles que celles de la perception. Une explication avancée pour l'augmentation de la durée des fixations est que le décodage d'une image mentale est plus complexe que l'encodage perceptif.

Une étude de Laeng et Teodorescu [Laeng02] réplique le paradigme de Brandt et Stark pour montrer que les sujets qui fixent le centre de l'image pendant la phase de perception adoptent spontanément la même position de regard pendant la remémoration. Pour ces mêmes sujets, lorsqu'il leur est demandé de bouger les yeux pendant la phase de remémoration, leur capacité à décrire la scène semble diminuer. Les mouvements oculaires ont ainsi un rôle d'index spatial permettant de décoder les données mnésiques. Cependant, cette baisse de performance mnésique pourrait aussi être due à un processus de surcharge attentionnelle causée par la tâche secondaire.

Malgré l'intérêt de ces approches, il est nécessaire de noter deux limites. Tout d'abord ces observations ne semblent se vérifier que pour des scènes visuelles très simples de type grille comportant des carrés noirs et blancs. La complexité de la scène aurait donc un effet sur ces processus. De plus, il est possible de catégoriser une scène visuelle en une seule fixation (c'est-à-dire

en 300ms) tout en se rappelant *a posteriori* le contenu de la scène [Henderson03c], la capacité de mémorisation n'est donc pas directement associée aux mouvements des yeux.

1.2.3.2 Le regard est synchronisé à l'activité motrice

Le regard obéit à un processus moteur permettant d'acquérir quasi instantanément de nombreuses informations sur l'environnement. Il est donc naturellement associé à l'action principale. Cependant quel type de synchronisation existe-t-il entre les gestes et le regard ? Mary Hayhoe est l'un des principaux chercheurs de ce domaine, étudiant les mouvements oculaires dans un contexte écologique. Hayhoe [Hayhoe00] enregistre les mouvements des yeux de participants à qui il est demandé de préparer des sandwichs dans une cuisine. Les résultats montrent que sur les 250 fixations enregistrées, moins d'une dizaine de fixations ont lieu sur des éléments visuels non pertinents pour la tâche. Dans une autre étude, [Hayhoe03] l'auteur demande de nouveau à des participants de préparer un sandwich et montre que l'enchaînement des actions motrices reproduit la séquence des fixations oculaires, le regard étant en avance d'une étape ; ainsi, tandis que la main prend le couteau, le regard s'oriente vers la confiture, puis l'autre main prend la confiture tandis que le regard s'oriente vers le pain, etc. Le regard permettrait donc le déploiement de l'attention afin d'anticiper spatialement l'exécution des gestes et synchroniser les différents mouvements.

1.2.3.3 Le regard est synchronisé à l'activité cognitive de lecture

Historiquement, la lecture est l'une des premières activités exploitant la mesure du mouvement des yeux [Baccino02]. En effet, du fait de la taille des lettres et des mots, l'activité attentionnelle est nécessairement locale, et du fait de la construction du sens, cette même attention est très fortement séquentielle. Ainsi, la position de la fovéa et l'extraction des informations visuelles sont bien corrélées. Rayner [Rayner98] propose un état de l'art des connaissances dans l'étude de la lecture et du traitement de l'information visuelle. Il recense trois types de processus oculaires constants lors de la lecture (en langue anglaise). Premièrement, les fixations durent environ 200ms avec des distances de saccade d'environ 7 à 9 lettres. Deuxièmement, la durée des fixations augmente et l'amplitude des saccades diminue lorsque le texte devient plus difficile à conceptualiser [Jacobson79]. Troisièmement, les mouvements oculaires diffèrent selon qu'il s'agisse d'une lecture à voix haute ou silencieuse. Carpenter [Carpenter83] montre que la probabilité qu'un mot spécifique soit fixé dépend de sa nature: 85% des mots avec un sens sémantique sont fixés contre 35% des mots fonctionnels (articles, adverbes, etc.). Cette probabilité de fixation semble influencée par la taille du mot : un mot d'une longueur 2 à 3 lettres a 75% de chance d'être survolé mais un mot de 8 lettres est quasi-systématiquement fixé. Les mouvements des yeux dépendent aussi de la lisibilité du texte [Morrison81], et des difficultés syntaxiques [Ferreira86]. Enfin, des différences de mouvements oculaires sont observées selon le niveau d'étude des participants [Dixon48].

1.2.3.4 Le regard est synchronisé aux processus du langage

Les relations entre les processus linguistiques et la perception visuelle sont intensivement étudiées depuis plusieurs années (pour une entrée bibliographique, voir [Henderson04a]). Quelques unes des observations fondamentales sont les suivantes. Tout d'abord, il apparaît une forte interconnexion en temps réel entre une instruction et le mouvement oculaire. Ainsi, lorsqu'il existe dans la scène deux objets phonologiquement proches, le regard se dirige suite à l'instruction en premier vers l'objet le plus probable en termes de phonème, puis lorsque le mot entier a été entendu, le regard se dirige vers le bon objet. Nous ne détaillerons pas ici ces relations complexes entre regard et linguistique mais nous retiendrons que la nature langagière de consigne dans une expérimentation ou un amorçage sémantique dans une tâche visuelle influera directement sur les mouvements des yeux.

1.2.3.5 Le regard est synchronisé aux processus d'échanges humain-système

Lorsqu'un utilisateur est exposé à une interface complexe, de nombreux flux d'informations doivent être appréhendés, analysés et évalués afin de développer une stratégie optimisée de recherche d'informations. Puisque l'utilisation d'une interface est en général associée à une tâche spécifique, il est considéré que l'attention est plutôt de nature sérielle. De plus, l'information est principalement exposée sous forme de texte et de graphiques de taille relative adaptée au regard. Le focus de l'attention visuelle est donc principalement fovéal de par le détail des indices visuels. C'est ainsi que la mesure du mouvement des yeux est utilisée en tant que pointeur de l'activité cognitive lors de l'exploration d'interfaces multimédia telles que les pages web, combinant image, texte, et structure complexe de navigation (site web) [Baccino01]. En ergonomie, les activités de conduite peuvent être étudiées en utilisant l'oculométrie ([Goldberg98], [Jacob03], [Liu01], [Donnet03]). Dès 1940 [Martin40], montre que la direction du regard peut expliquer des erreurs de pilotage d'avions, et que la fréquence des saccades permet de mesurer la fatigue de conduite automobile sur une longue période [Specht41]. Toujours dans les tâches de conduites, la mesure oculométrique a permis de montrer que l'orientation du regard à l'approche d'une courbe est corrélée à l'angle du virage [Land94]. La comparaison des regards de pilotes confirmés et de novices permet de mettre en évidence des stratégies visuelles d'expertise face à une interface complexe [Land01]. Enfin, Simon [Simon93] a quant à lui évalué la charge de travail des opérateurs à partir de leurs mouvements oculaires et obtient des résultats très satisfaisants.

1.2.3.6 Le regard est synchronisé aux processus d'évaluation visuelle: le cas de l'évaluation du design

Dès le début des années 1990, des chercheurs s'intéressent à l'apport de l'oculométrie dans la compréhension des processus d'évaluation de produits de consommation. Ainsi, Hammer et Langyel [Hammer91] effectuent des études de mouvements oculaires sur des objets manufacturés. Leur but était de comprendre l'apport de cette méthode pour l'aide à la conception à destination des designers. Ils observèrent les zones des objets (appareils électroménagers et automobiles) attirant l'attention visuelle. Le logo et le nom de la marque apparaissent comme les principales zones traitées consensuellement entre tous les participants.

Un groupe de travail finlandais [Koivunen04] a étudié la manière dont les participants évaluent le design d'objets manufacturés (plus spécifiquement des téléphones mobiles). L'analyse des mouvements oculaires montrent que le même objet (produit de consommation) n'est pas exploré de la même manière selon la tâche. La typologie des mouvements fournirait trois classes de stratégies exploratoires : spatiale, par exploration de proche en proche (« narrow »), holistique c'est-à-dire par niveau sémantique du contenu (« holistic »), ou enfin par combinaison des deux premières stratégies. Sharmin et collaborateurs [Sharmin04] en collaboration avec Nokia, approfondissent les méthodes d'évaluation du design. L'approche consiste à considérer que les mouvements oculaires sont bien connectés aux processus cognitifs mais que la mesure de la direction du regard en elle-même n'est pas suffisante. Il serait ainsi pertinent d'observer en complément la connaissance préalable de l'objet par le participant et ses intentions afin d'interpréter son comportement oculaire. Cette équipe combine donc la mesure oculométrique et des méthodes complémentaires afin de distinguer les effets de saillances physique (couleur, forme...) des effets de prégnance cognitive (l'information sémantique). Le paradigme employé consiste à mesurer le mouvement des yeux lorsque des participants évaluent sur photos des objets tels qu'appareil photo, téléphone, tasse, ou cisailles. Les consignes, exécutées par des groupes indépendants, sont les suivantes : tâche de première impression, de mémorisation, évaluation esthétique, évaluation d'usage et de durabilité. Les résultats montrent que selon la consigne, la durée des fixations varie (consigne de première impression équivaut à des fixations plus longues ; esthétique à des fixations plus courtes). De plus, un objet simple produira moins de fixations

qu'un objet complexe (exemple : tasse *versus* tondeuse). Dans cette même étude, la comparaison expert en conception du design / novices (10 sujets par catégorie) montre un effet de l'expertise sur le comportement visuel pour la consigne de mémorisation mais l'analyse des variables indépendantes comme la profession, l'expérience, le genre, le type de loisirs, ou le degré de familiarité avec le produit, ne montre pas d'effets de ces facteurs sur la prise d'information visuelle. Le regard peut donc être utilisé en tant que trace du traitement de l'information visuelle, mais ces analyses nécessitent de grandes précautions car de nombreux facteurs dirigent l'attention visuelle.

1.2.3.7 Le regard démontre l'expertise dans des tâches de diagnostic

On aborde ici l'expertise diagnostique, c'est-à-dire la capacité à traiter visuellement la scène de manière optimisée. Ainsi, les mouvements oculaires peuvent être utilisés pour analyser la manière dont un expert effectue une tâche d'inspection. Schoonard, Gould et Miller [Schoonard73] observent en comparant experts à la tâche d'inspection et novices que les inspections efficaces de cockpit d'avion pour le diagnostic de défauts sont caractérisées par une bonne précision de la position du regard, par de brèves fixations et des saccades rapides. De même, Megaw et Richardson [Megaw79] observent un seuil de durée de fixation au delà duquel une cible recherchée a été détectée dans la zone fixée. On relie ainsi le comportement oculomoteur à l'efficacité de la recherche de cible, ainsi qu'à sa découverte. Duchowski et collaborateurs [Duchowski01] et Wang & Lin [Wang97] ont quant à eux utilisé l'eye-tracking en tant que système permettant d'apprendre à des participants à effectuer des explorations visuelles efficaces lors de la vérification visuelle d'une carlingue d'avion. Après chaque inspection, des indications sont fournies afin de modifier le trajet du regard à la prochaine inspection. Une expérience en environnement virtuel immersif a été menée. Elle consistait à équiper les participants d'un système d'écrans porté sur tête (HMD) et d'enregistrer les mouvements des yeux pour la détection de la corrosion dans une cabine d'avion. Lorsque les sujets échouaient, leurs mouvements oculaires étaient analysés et des indications de correction étaient fournies [Bisseret99].

Law, Atkins, Kirkpatrick, Lomax, et Mackenzie [Law04] mesurent la vitesse et la précision des mouvements oculaires au cours de l'entraînement à des actes chirurgicaux, en particulier des laparoscopies. La comparaison des mouvements des yeux de chirurgiens experts et de novices dans ces interventions montre que le comportement du regard est révélateur du niveau d'expertise du participant, et donc que la prise d'information est corrélée aux connaissances du participant. Ces résultats montrent que dans un contexte de diagnostic, le regard est un bon indicateur de la recherche d'information visuelle, et qu'il manifeste le niveau d'expertise à la tâche du participant.

1.2.3.8 Le regard est synchronisé avec les processus de prise de décision

Ce champ de recherche récent a principalement été étudié dans le laboratoire de Shinsuke Shimojo à l'université de Californie notamment par Simion & Scheier [Shimojo03]. La proposition de base consiste à considérer que les processus d'orientation du regard et de prise de décision sont intrinsèquement liés à l'attitude et l'attention du participant envers un objet de scène. Ainsi, l'orientation du regard témoignerait de l'attitude du participant envers l'objet (en psychologie l'attitude est une manifestation affective orientée, telle que l'attractivité ou le rejet) puis participerait à la prise de décision. Pour vérifier ces hypothèses, les auteurs emploient dans un premier temps un paradigme de présentation par paire sur des images virtuelles de visages. Plusieurs consignes sont définies : lequel aimez-vous le plus ? Lequel aimez-vous le moins ? Lequel est le plus rond ? Le participant signifie lui-même la fin de l'essai en appuyant sur un bouton lorsque son choix est effectué. En enregistrant les mouvements du regard, il apparaît que dans la seconde qui précède la prise de décision (appui sur le bouton), le taux de regard sur l'image qui va être choisie augmente de manière significative jusqu'à la décision, pour atteindre un taux compris entre 70 et 90% quelle que soit la nature et la difficulté de la tâche. La typologie de la courbe de ce taux dans la dernière seconde des essais donne le nom à l'effet « gaze cascade ». La deuxième observation de ces auteurs est que la

tâche d'évaluation de la préférence est spécifique, alors que les tâches de rejet et d'évaluation géométrique sont plutôt similaires. Il semble apparaître une interaction entre l'orientation du regard et la décision produisant une orientation très significative du regard en direction de l'image préférée. Faisant appel au « mere exposure effect » (effet d'exposition seul) [Zajonc68], ils proposent pour la consigne de préférence que plus l'image est observée, plus elle est appréciée, et plus elle est appréciée, plus elle est observée.

1.2.3.9 Synthèse

Le regard est un indicateur complet et complexe de l'activité attentionnelle, mnésiques ou de prise de décision. Les situations où le regard révèle l'activité cognitive sont très variées ; cependant du fait de la multiplicité des facteurs interagissant dans le guidage du regard, les interprétations doivent être effectuées avec prudence. Alors que les paradigmes exposés précédemment évoquaient des situations dans lesquelles les mouvements des yeux étaient de nature à encoder des informations visuelles, nous nous intéressons dans le paragraphe suivant à l'usage des systèmes oculométriques dans des situations où le regard a pour fonction d'indexer un élément de la scène visuelle.

1.2.4 Paradigmes d'études des mouvements oculaires pour l'indexation des informations de scène

1.2.4.1 Le regard utilisé en tant que pointeur interactif pour les interfaces homme-machine

Deux postulats sont posés pour l'utilisation du regard dans un contexte d'interface homme machine. Le premier considère que la position du regard et le niveau d'attention dans la scène sont corrélées. Le deuxième propose que dans une situation de manipulation d'objet, l'action et le regard soient corrélés et contrôlés. Des systèmes ont ainsi été développés afin de faire interagir la position du regard sur un écran et le contenu exposé par cet écran. La souris d'ordinateur est par exemple remplacée par ce système pour les personnes à déficience motrice⁵. Un simple entraînement de quelques heures suffit à maîtriser le déplacement du pointeur par le regard, à sélectionner des icônes, et ainsi à interagir avec un ordinateur [Stampe95].

Le système « eye-typing » est construit sur ce principe. Il consiste à écrire des mots en fixant des lettres sur un clavier virtuel [Majaranta02]. De même, une application d'assistance par le regard a été développée par Hyrskykari et collaborateurs en 2003 [Hyrskykar03] pour l'aide à la lecture des documents électroniques en langue étrangère. Pendant la lecture des documents, le système « iDict » affiche contextuellement des informations aidant à la compréhension du texte en fonction de la phrase ou du mot observé.

Qvarfordt et Zhai [Qvarfordt05] proposent dans une étude de 2005 menée en partenariat avec IBM [Qvarfordt05] un système d'interface homme machine basé sur les mouvements oculaires. Lorsque le participant observe un plan d'une ville sur un écran, le système « iTourist » analyse en temps réel la position du regard, la durée depuis laquelle il est dans la même zone, la fréquence d'entrée et sortie sur la zone, la relation catégorielle entre les objets de la zone et ceux de la zone précédemment observée et la distance de saccade effectuée pour parvenir à la zone. En fonction de ces paramètres, le système tente de prédire la zone d'intérêt active de l'utilisateur. Des informations vocales adaptées sont alors fournies. 12 participants testent ce système ; tous réussissent l'ensemble des tâches qui

⁵ De nombreuses informations sont disponibles sur le site web suivant : <http://www.cogain.org>

leur était demandé alors qu'elles nécessitaient la prise en compte des informations auditives activées par le regard.

Un autre système technologique utilisant la position du regard a été développé par Canon en 1995 sur un appareil photo (UC-X1 Hi) qui intégrait une fonction d'« eye-tracking ». L'autofocus de l'appareil était basé sur la direction du regard. Il suffisait de regarder une zone pour que la netteté s'effectue dessus. Un manque de fiabilité était cependant présent par des pertes brutales et aléatoires du « tracking », obligeant l'utilisateur à effectuer un mouvement brusque de l'œil pour re-calibrer le système [Hifi & elektronik; 1995] et un petit carré blanc affiché en permanence sur le viseur au niveau de la fovéa gênait la vision naturelle.

Ainsi, diverses applications ont d'ors et déjà été trouvées à l'utilisation des oculomètres en tant qu'interface homme-machine. L'augmentation de la puissance des unités de calcul associée à la miniaturisation des systèmes électroniques devraient permettre un fort déploiement de ces techniques dans les années à venir.

1.2.4.2 La position du regard exploitée pour l'optimisation du rendu des images de synthèse

Les systèmes de « gaze contingent display » se basent sur le caractère non homogène de la résolution de la vision sur l'ensemble du champ visuel, en considérant que seule la zone de la scène sur laquelle se porte la fovéa est analysée à un haut niveau de détail. Le niveau de résolution de l'information affiché sur un écran est ainsi rendu en temps réel, à haute résolution dans la zone fovéale et de manière dégradée dans les zones périphériques du champ visuel. Ce type de système permet d'optimiser le temps de calcul des moteurs de rendu temps réel d'image de synthèse dans des environnements interactifs, et par exemple dans une situation d'évaluation diagnostique, permet que le spectateur ait la même vision de la scène en termes de résolution que l'expert qui effectue la tâche [Dorr05]. Hormis la problématique du temps réel, cette approche permet de transformer des photos en images hybrides photo-réalistes dont les zones contenant des informations pertinentes sont rendues à haute résolution et les autres zones de manière dépeçurée (« non photorealistic rendering » ; [Barth06], [Santella02]).

1.2.4.3 Le regard utilisé dans les systèmes interactifs multimodaux

Le regard peut aussi être utilisé comme régulateur de situations ambiguës en communication. Dans un environnement virtuel, par exemple de télé conférence multi-participants, le regard peut servir de pointeur afin de désigner l'avatar (la représentation virtuelle d'un individu) vers qui diriger un flux audio (parole par exemple) [Zhang04]. De la même manière, la mesure de la position du regard peut permettre de placer un indicateur sur un document électronique indiquant la zone qu'un participant est en train de lire. Cet indicateur est reproduit sur le système de l'interlocuteur, indiquant la zone qui est par exemple évoquée verbalement [Vertegaal99].

Les travaux de Stefan Raidt [Raidt07], consistent ainsi à utiliser des oculomètres croisés afin de mesurer la direction du regard de chacun de deux interlocuteurs discutant par écran vidéo interposé. Il est alors possible d'effectuer un rendu en temps réel de la direction du regard d'un avatar représentant notre interlocuteur, la communication s'en trouvant remarquablement optimisée.

1.2.5 Synthèse

On a pu voir dans cette partie consacrée à l'utilisation des systèmes oculométriques que de nombreux usages peuvent découler de l'exploitation du rôle de pointeur et du rôle de capteur que revêt le regard. Tandis que l'étude des mouvements d'encodage est pertinente pour l'analyse des processus de perception visuelle et l'inférence des processus cognitifs, l'utilisation des mouvements

d'indexation permet d'élaborer de nouvelles manières d'interagir avec un système, et d'exploiter la mesure en temps réel de la zone du champ visuel vers laquelle l'attention s'oriente.

Après avoir présenté de nombreuses situations exploitant les techniques de mesure de la position du regard et s'être appuyé sur des champs théoriques très différents, il est nécessaire de positionner le cadre général de notre étude en s'intéressant aux processus associés aux paradigmes que nous manipulons. C'est ainsi que la partie suivante permet de préciser le concept de scène et d'objet de scène qui seront exploités dans notre étude.

Chapitre 2

Scène, information et vision

L'objectif de notre travail est d'établir des liens entre la mesure des mouvements oculaires, les processus perceptifs et attentionnels, et l'évaluation visuelle d'un objet manufacturé. Dans ce chapitre, nous proposons de spécifier le type de scène que l'on va traiter, en étudiant les différents niveaux informationnels contenus par une scène, et les relations entre une scène et les objets qu'elle contient.

2.1 De la scène à sa représentation dépeinte

2.1.1 Définition du concept de « scène »

Dans le but de classer les scènes visuelles selon leur nature, il est pertinent de reprendre la définition de base de ce mot, afin d'en dégager les propriétés principales. Selon le dictionnaire Meryam-Webster, une scène est une « zone d'arrière plan d'un jeu théâtral dramatique, l'une des subdivisions d'une zone de jeu, la zone d'une activité ou action ». Nous tirons de ces définitions plusieurs indices sur le concept de scène. Tout d'abord, une scène est une zone, donc une partie de l'espace délimité par des frontières. Il s'agit du tout ou d'une partie du champ visuel. Le cadre est l'espace occupé par la scène dans le champ visuel, par opposition au hors cadre, tout ce qui n'est pas dans l'espace de la scène. Il existe donc des frontières à la scène visuelle, délimitant le cadre. Ces frontières sont des limites perceptives : contours, transitions de textures, contrastes, mais aussi des limites attentionnelles : ce sur quoi ne peut porter l'attention dans un processus cognitif continu. Le cadre peut être lié au hors cadre par une relation de trans-ordinalité (contiguïté physique), ou de super-ordinalité (inclusion de la scène dans son contexte). Le concept de scène est de plus associé à l'idée d'action. La scène doit donc obéir à un continuum d'espace et d'action, à partir duquel le hors cadre peut aussi être défini.

Ensuite, une scène est composée de plusieurs plans, à différentes profondeurs, du premier plan à l'arrière plan. Cela implique l'idée de « point de vue d'observation » car la profondeur (« avant, arrière ») ne se définit que depuis un point, depuis lequel un plan occulte un autre, indiquant leur proximité relative. Chaque plan est défini par une répartition d'objets à la même profondeur, donc subissant la même perspective.

D'un point de vue théorique, la scène est donc un ensemble de plans formés par des objets, possédant une homogénéité spatiale et temporelle, depuis et vers laquelle s'effectuent des échanges d'informations avec un point de vue extérieur à la scène.

Les expériences portant sur la perception visuelle et plus particulièrement sur les mouvements oculaires utilisent des stimuli qui peuvent être classés en plusieurs catégories en fonction de leurs niveaux de représentation. Ainsi, certaines études portent sur des objets en tant que formes géométriques simples, c'est-à-dire des représentations symboliques ; il s'agit selon Chauvin [Chauvin03] d'une classe de percepts se basant sur un matériel verbal (« cercle », « carré »). D'autres études utilisent des scènes abstraites, par exemple des formes de couleur ou des fractales.

Cette classe permet d'étudier le comportement visuel sans l'existence de dimensions sémantiques (bien que naturellement il persiste une activité de création de sens). La troisième classe est celle des matériels écologiques, toujours selon Chauvin, correspondant à la réalité, que nous subdiviserons en 2 sous-classes, celle de la réalité et celle des représentations photo-réalistes. Nous nous intéressons dans le paragraphe suivant aux différences entre une scène réelle et sa représentation, dite « dépeinte » ('depicted').

2.1.2 Différences entre scène réelle et scène dépeinte

Il est effectué couramment la distinction entre scène du monde réel et scène dépeinte. La scène réelle est celle que le lecteur a devant les yeux à l'instant tandis que la scène dépeinte est une représentation photographique, peinte, ou virtuelle de cette scène.

Les scènes réelles et dépeintes se différencient par plusieurs caractéristiques :

- La surface du champ visuel est différente, car les scènes réelles remplissent l'ensemble du champ visuel tandis que les scènes dépeintes sont délimitées par un cadre.
- Les scènes réelles produisent un spectre de longueur d'ondes et d'intensité lumineuse dont la dynamique est supérieure à celle des scènes dépeintes.
- Les scènes réelles suivent des règles sémantiques et syntaxiques spécifiques (inclusions, superpositions, lois physiques), dont peut s'affranchir la scène dépeinte.
- La perception de la profondeur de champ est absente des scènes dépeintes.
- Le nombre d'entités visuelles est plus élevé dans la scène réelle que dans la scène dépeinte.
- Le point de vue est unique sur la scène dépeinte tandis qu'il peut évoluer dans la scène réelle
- L'implication du participant (croyance dans le médium) est supérieure pour une scène réelle que dépeinte

On constate donc que la quantité d'informations est fortement réduite lorsque l'on passe de la scène réelle à la version dépeinte de cette même scène. Quelles conséquences cela a-t-il sur le comportement visuel ?

2.1.3 Un comportement visuel propre à la nature réelle ou dépeinte de la scène

Des expérimentations sur le monde réel telles que celles de Land et Hayhoe [Land01b], étudiant le comportement visuelle dans des activités de la vie quotidienne ou celles de Land et Lee [Land94], lors d'activités de conduite, montrent qu'il existe des différences de comportement oculomoteur entre les observations en situations du monde réel et sur images. Une étude comparative de Tatler et collaborateurs [Tatler05b] confirment que les informations visuelles (position d'objets, couleurs, formes...) ne sont pas traitées de la même manière selon que la scène soit réelle ou dépeinte, certaines informations nécessitant plusieurs fixations pour être encodées dans la scène dépeinte, et une seule dans le monde réel, ou inversement. Ainsi, selon la nature de la scène, certaines informations sont encodées temporairement et d'autres sur le long terme. Henderson et Hollingworth [Henderson98] et Land et Hayhoe [Land01b] montrent que les amplitudes de saccades diffèrent selon la nature de la scène. Sur image, les amplitudes sont en moyenne de 1 degré et dépassent rarement 10 degrés, tandis que dans le monde réel, les saccades sont rarement inférieures à 5 degrés et montent jusqu'à 20 degrés. Les différences peuvent provenir du facteur d'échelle, ou encore du grand nombre d'items dans la scène réelle dû à la plus grande dimension du champ visuel (180 vs 40 degrés

environ). Tatler observe que les durées de fixations moyennes sont plus importantes pour la même tâche sur la scène réelle que sur la scène dépeinte. Ces éléments confirment le fait que le traitement perceptif et attentionnel est dépendant de la nature du support.

Notons qu'un effet oculomoteur spécifique a été observé lors d'études de scène sur des écrans, et qu'il est absent dans le cas d'études sur des scènes réelles. Il s'agit du biais d'observation central ([Parkhurst02], [Vincent07]), le regard se posant naturellement et préférentiellement sur la zone centrale de la scène.

Malgré tout cela, la majorité des paradigmes employés pour l'étude de la perception visuelle sur les scènes réelles utilise des représentations de scènes, c'est-à-dire des images, afin de simplifier l'étude. L'utilisation des images en tant que modèle proche de la réalité se base sur le postulat que les propriétés importantes de la scène telles que la complexité visuelle, les informations sémantiques et structurales ainsi que le sens global, sont identiques dans la scène réelle et dans la scène dépeinte. Il s'avère de plus que les différences entre scène réelles et scènes dépeintes semblent moins limitatives que les difficultés expérimentales rencontrées dans le contrôle des différents facteurs dans le monde réel. Pour ce travail, nous nous situons dans le cas d'études de scènes dépeintes, autrement dit de représentations photo-réalistes de scènes réelles.

2.2 Propriétés des scènes dépeintes

Alors que l'on a vu dans la section précédente que les scènes réelles et scènes dépeintes contiennent des informations différentes, on va s'intéresser aux dimensions qui distinguent les différentes scènes dépeintes, abstraites ou concrètes, stylisées ou réalistes, complexes ou simples.

2.2.1 Deux dimensions orthogonales : abstraction/concrétisation et stylisation/réalisme

Il est possible de classer les différentes images de scènes selon deux dimensions, la première dimension décrivant la forme perceptive de l'objet visuel, et la deuxième dimension le type de sémantique activé par ce même objet.

La forme perceptive peut être définie sur un continuum, dont une extrémité est la scène réaliste, et dont l'autre extrémité est la scène stylisée. La scène réaliste contient un grand nombre d'informations de luminance, de chrominance, produisant une impression de réalité. La scène stylisée est une représentation de scène simplifiée, par exemple noir et blanc, dessinée au trait. Différentes études en perception visuelle utilisent des représentations positionnées sur cet axe : dessins au trait, représentant des scènes concrètes [Loftus78] ou des scènes dessinées sur la base de photographies [DeGræf90], photographies couleurs [Henderson99], images de synthèse [Duchowski01], rendu non photo-réaliste de photographies [Santella02], photographies filtrées en fréquences [Oliva97]. La forme symbolique peut être définie sur un continuum dont une extrémité est la scène abstraite et l'autre extrémité est la scène concrète. Une scène abstraite est par exemple une image fractale. Une scène concrète peut être une photographie ou encore une image de synthèse d'un jeu vidéo.

La classe des scènes concrètes peut être subdivisée en sous-classes, selon les dimensions de naturalité et artificialité.

2.2.2 Critères de distinction des scènes naturelles/artificielles

Des observations neurocognitives du traitement de l'information visuelle montrent qu'il existe des processus spécifiques au traitement des scènes « naturelles [Guyader04], pour lesquelles l'être humain est censé être spécifiquement adapté, contrairement aux scènes « artificielles ». Classifier les scènes visuelles réalistes selon leur niveau de naturalité semble ainsi pertinent. Il existe au moins deux critères de discrimination des scènes : scène d'intérieur/scène d'extérieur et scène de ville/scène de paysages. La propriété « fait par l'homme » *versus* « naturel » pourrait ainsi activer des processus de traitement de l'information visuelle spécifiques, où le traitement des scènes naturelles serait plus rapide et efficace que celui des scènes artificielles. Afin de classer les différents types de scènes, Henderson [Henderson04a] propose une projection en deux dimensions : dimension « scène réelle/dépeinte » *versus* dimension « organisation de scène naturelle/artificielle ». Cependant il existe une nuance, selon que l'on traite d'objets ou de scènes : le degré de naturalité serait, lorsque l'on s'intéresse à des objets, la manière dont ceux-ci sont disposés dans la scène, s'il s'agit d'une organisation artificielle ou non. Lorsque l'on traite de scènes, la dimension de naturalité renverrait à la représentation : s'agit-il de scènes d'intérieur ou de paysages. Cette approche atteint cependant sa limite lorsqu'il s'agit de définir le degré de naturalité d'une photographie d'une maison dans les bois, car il s'agit d'une scène naturelle avec un élément artificiel. C'est pourquoi la distinction naturel/artificiel s'emploie plutôt pour des études sur des ensembles d'objets tandis que le degré d'ouverture de scène ('openess') [Guerin00] est utilisé pour différencier les scènes d'intérieur et scènes d'extérieur.

Notre travail va porter sur des photographies d'objets ; il s'agit donc d'une scène dépeinte concrète, représentation réaliste du monde réel dont les couleurs auront été étalonnés. Selon Henderson et Hollingworth [Henderson 99b], les scènes photo réalistes répondent à plusieurs critères : il s'agit d'un ensemble sémantique cohérent, cet ensemble doit être nommable, et la métrique à échelle humaine, les lois physiques contraignent l'ensemble (respect de la gravité, pas de superposition spatiale, occlusions par perspectives), les lois sémantiques structurent l'ensemble, c'est-à-dire qu'un volant n'est pas disposé sur un siège, ou une télévision dans un four.

2.2.3 Définir la complexité des scènes visuelles

La complexité d'une scène renvoie à la densité d'objets présents ainsi qu'à leur cohérence. Plus une scène est complexe, plus les processus attentionnels auront tendance à traiter de manière séquentielle les informations visuelles et plus longue sera la phase d'exploration et d'identification. Ainsi, il peut être pertinent de définir un niveau de complexité des scènes. Malgré le fait que les scènes réelles soient composées de nombreux objets, textures et couleurs, nous sommes capables d'analyser très rapidement l'essentiel de la scène.

Heaps et collaborateurs [Heaps99] proposent que la complexité soit corrélée à la difficulté à décrire verbalement l'image⁶. Plus le nombre de mots nécessaires à la description de la scène est élevé, plus la scène est complexe. La complexité se définirait alors sur un axe « du plus organisé et structuré » décrit en peu de mots (moins complexe), au « moins organisé » décrit avec beaucoup de mots (plus complexe). Van der Helm [van der Helm00] analyse la complexité perceptive de textures visuelles et propose (sans qu'il n'y ait d'arguments pour en interdire l'extension à d'autres propriétés visuelles) que les répétitions, symétries et similarités soient des critères permettant de classer les

⁶ On doit cependant distinguer deux processus : une scène peut être visuellement simple mais complexe à décrire (abstraction) ou complexe à visualiser (beaucoup d'éléments naturels) mais simple à résumer.

textures par groupes de complexité. Plus il y a de répétitions de motifs, plus ces motifs sont symétriques, et plus il y a de similarité entre eux, moins les textures sont considérées comme complexes. Selon Oliva [Oliva04b], la complexité d'une scène est liée à la variété des stimuli visuels, c'est-à-dire que plus le nombre d'entités locales est élevé dans la scène, plus la complexité est forte. Par entité est entendu tant des items sémantiques que des matériaux ou surfaces. Oliva [Oliva04b] propose et vérifie expérimentalement une définition multidimensionnelle de la complexité, qui serait tributaire de : la quantité d'objets de scène, la densité des objets, la variété des couleurs, le degré d'ouverture de scène, l'existence de symétries dans la scène et la configuration spatiale des objets dans la scène. Le niveau d'inclusion ou d'imbrication semble aussi influencer sur la complexité, c'est-à-dire que s'il est difficile d'identifier et séparer chaque entité de la scène, la complexité augmente. Enfin, les facteurs cognitifs tels que la pré-connaissance de la scène, de la situation, et de la nature de la consigne ont potentiellement une action sur le caractère de complexité d'une scène visuelle.

2.2.4 Synthèse

Les scènes photo-réalistes correspondent ainsi à un cas particulier de l'ensemble des scènes visuelles qui peuvent être exploitées dans les études en perception visuelle. Il est important de bien définir le type de scène (concrète/abstraite, intérieur (fermé)/extérieur (ouvert), disposition naturelle/artificielle) qui sera employé dans nos études car des processus perceptifs et cognitifs différents peuvent être activés selon le type de scène employé. La complexité des scènes est aussi une dimension qui doit être maîtrisée car elle impacte directement les processus attentionnels mis en œuvre lors de l'observation des scènes. Ce paragraphe a permis de voir qu'une scène est à la fois une entité propre, mais aussi la somme des éléments qui la constitue. On va donc s'intéresser dans les paragraphes suivants aux différents niveaux d'informations contenus dans une scène et aux relations qui lient la scène et les objets de scène.

2.3 De la scène aux objets qui la compose

La scène a été vue dans le paragraphe précédent comme une entité globale composée d'éléments locaux. Elle est construite sur la base de groupes perceptifs, mais peut être traitée à un niveau sémantique. Ainsi, lorsque l'on étudie les processus perceptifs et cognitifs de la vision, il est nécessaire de définir à quels types d'informations de scène on fait appel. C'est ainsi que le présent paragraphe va présenter la variété des niveaux d'informations contenus dans une scène, puis les différentes interprétations des signes visuels, et enfin les liens entre la scène et les objets qui la compose.

2.3.1 Structure holistique des informations de scène

Il existe plusieurs manières de définir les différents niveaux d'interprétation des informations contenues dans et par une scène. C'est ainsi que James et Chang [James00] proposent une modélisation pyramidale (Figure 7) des différents niveaux d'informations, en fonction de la quantité de mots nécessaires pour décrire l'image, modèle qui a été appliqué dans le cadre de l'indexation sémantique des images.

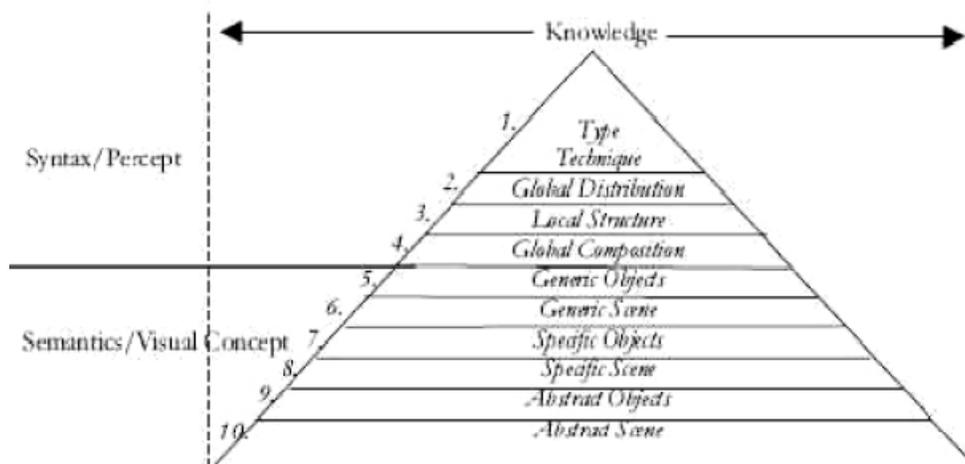


Figure 7 Pyramide de James & Chang [Jaimés00]

Cette pyramide comporte quatre niveaux perceptifs et six niveaux sémantiques, où la largeur de la base de chaque niveau est proportionnelle à la quantité d'information requise pour décrire le contenu de l'image à ce niveau. Les niveaux perceptifs définissent les éléments visuels tels que les couleurs et les lignes. Les niveaux sémantiques sont liés aux concepts et au sens des éléments visuels et de leur arrangement. Trois classes existent : générique, spécifique ou abstrait. Ainsi, une photo de voiture peut être une instance de la classe « véhicule » (générique), ou encore de la classe « Peugeot » (spécifique), ou enfin de la classe « luxe et confort » (abstrait).

Dans un cadre simplifié, Jorgensen [Jorgensen96] propose de classer les informations contenue par une image selon le niveau de traitement nécessaire à leur exploitation. Trois classes sont définies :

- information à visée perceptive (perception des couleurs, reconnaissance des formes)
- information à visée réactive (modification de l'état affectif : surprise, confusion, appréciation)
- information à usage interprétatif (émergence des symbolismes)

Une image peut donc être décrite par des niveaux d'information plus ou moins complexes, et l'intégration de ces informations par un individu est aussi plus ou moins élaborée. L'évaluation d'un produit par un individu est donc dépendante du fait que le produit peut être appréhendé à différents niveaux d'interprétation, et que l'évaluation sollicite différents niveaux de traitement de ces informations. Les scènes réalistes possèdent donc une structure informationnelle holistique, dont le niveau le plus élevé est celui de l'objet super-ordinal et dont le plus profond niveau est celui de l'objet sub-ordinal. Cette structure impose une cohérence spatiale et sémantique à l'ensemble. Afin de mieux comprendre les relations qui lient la scène et les objets de scène, nous nous intéressons dans un premier temps aux différentes définitions d'un objet.

2.3.2 Objets de scène, approche sémiotique

L'image d'un objet est à la fois une projection de la réalité et une réalité en soit. En effet, le contenu de l'image est une représentation de la réalité, dépeinte sur un support, mais le support de l'image est réel pour l'observateur. Il existe donc plusieurs niveaux d'informations :

- l'information délivrée par un capteur est un ensemble de signaux variant en amplitude et phase

- l'information délivrée par un algorithme de reconnaissance de forme est un ensemble d'entités identifiées par une homogénéité intrinsèque de texture, couleur, contours et contrastes.
- l'information délivrée par une description verbale est un ensemble d'objets ou propriétés sémantiques.

L'image est donc intégrée en prenant en compte un ensemble d'informations codées de différentes manières à différents niveaux. L'approche sémiotique propose que l'image soit composée de signes, tout objet étant un signe, où le signe transcende l'objet. Nous nous intéressons ainsi à la construction du signe.

« Un signe est tout ce qui peut être, dont le but est de médiatiser entre son émetteur et son interprète, tous deux dépositaires de la pensée, par la transmission d'une signification du premier au second » [Pierce1907]. Cette définition amène l'idée que le signe est une liaison entre un percept et une interprétation. On peut cependant aller plus loin dans la décomposition de cette liaison.

La morphologie triadique du signe est proposée par Pierce qui distingue trois types de signes : l'icône, l'indice et le symbole (Figure 8).

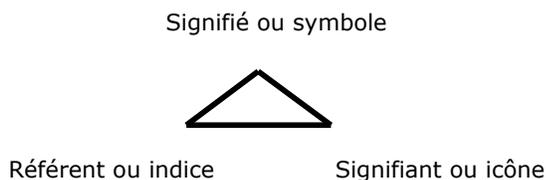


Figure 8 Relation triadique de Pierce [Pierce1907]

- Le signifiant (ce qui est perçu) ou icône, est la représentation perceptive d'un objet ;
- Le signifié (la signification attribuée), ou symbole, est le signe conventionnellement associé à son objet, et perdrait le caractère qui ferait de lui un signe s'il n'y avait pas d'interprétant ;
- Le réfèrent (ou indice) est ce que le signe représente dans sa réalité physique ou conceptuelle, la nature physique de l'objet ou une corrélation à sa nature.

Laudati [Laudati00] propose ainsi «qu'un signe est quelque chose (signifiant) qui tient lieu pour quelqu'un de quelque chose (réfèrent), sous quelque rapport ou à quelque titre (signifié)». Le sens peut donc être tiré de l'image prise dans sa globalité ou de chacun des signes de l'image pris distinctement. Ces différentes lectures ne convergent pas nécessairement.

2.3.3 Objets de scène, approche cognitive

Les objets sont classiquement définis en psychologie cognitive par leurs traits et leurs propriétés. Les propriétés d'un objet correspondent à sa représentation conceptuelle et à ses fonctions tandis que les traits d'un objet sont les éléments permettant de décrire l'objet perceptivement. Cordier et Tijus (cités par [Domnrier06]) proposent de décrire les objets par 8 caractéristiques :

- Les traits perceptifs encodent les informations telles que la forme, la couleur, et la texture ;
- Les traits structuraux décrivent l'organisation configurale des parties d'objets pour former le tout. La relation holistique de ces parties par imbrication est ici fondamentale, par exemple un habitacle est une partie de la voiture mais est composé d'un volant et

de sièges, eux-mêmes des objets. Les traits structuraux définissent aussi les matériaux, la quantité et la répétition des parties, ainsi que la situation ;

- Les traits fonctionnels, sont eux liés à l'usage de l'objet, et qui pour les objets manufacturés déterminent les propriétés structurales ;
- Les traits procéduraux sont les différents usages de l'objet, mettant en œuvre les différentes fonctions, et sont liés à l'action ;
- Les traits comportementaux décrivent les manières d'agir des êtres vivants, c'est-à-dire des activités ;
- Les traits de personnalité sont les catégories biologiques auquel un état psychologique est attribuable pour un être humain ;
- Les traits lexicaux encodent quant à eux les noms associés à l'objet et concernant la catégorie à laquelle il appartient : exemple : une Peugeot, un ordinateur ;

Dans le cadre de notre travail, nous nous intéressons aux traits perceptifs, structuraux et lexicaux de l'objet étudié vis-à-vis des zones visuelles traitées par le regard.

2.3.4 Objets de scène, approche enactive

On peut définir que les objets de la scène sont les éléments permettant une affordance humaine ainsi qu'une interaction motrice ; Gibson propose en effet que les objets sont définis par « toutes les actions possibles latentes dans l'environnement, mesurables objectivement et indépendantes des capacités individuelles à les reconnaître » [Gibson77]. Selon cette définition, un objet est ce vers quoi on interagit, et donc ne se décompose pas en sous-objets. Un objet est donc à échelle et portée de manipulation humaine. A partir de cette approche on peut considérer qu'une scène diffère d'un objet car le point de vue d'observation de la scène peut être inclus dans la scène tandis que l'observation d'un objet est nécessairement effectuée depuis l'extérieur. L'affordance devient ainsi un critère de distinction scène/objet.

2.3.5 Objets de scène, approche neuro-perceptive

Engel [Engel97] montre que les scènes et objets de scène ne sont pas traités par les mêmes processus neurocognitifs. Les informations visuelles qui proviennent de la fovéa et celles provenant des zones para-fovéales sont traitées par différentes zones du cortex visuel où la majorité du traitement est orientée vers les informations fovéales. Ainsi, la fovéa est orientée vers les zones visuelles qui nécessitent un traitement approfondi tandis que les zones qui nécessitent un large champ spatial pour être traitées le sont par la vision périphérique. Donc, selon que l'attention s'oriente vers un objet ou la scène globale, différentes zones corticales seront activées. Notons les travaux d'Epstein et Kanwisher [Epstein98] : en utilisant l'Imagerie par Résonance Magnétique, ils observent que le cortex temporal inférieur est activé lorsque les participants observent des photographies de scènes réelles d'intérieur ou d'extérieur d'habitations, ce qui n'est pas le cas lorsque les participants observent des visages ou des alignements d'objets. La différence entre ces deux groupes de scènes pourrait être la taille du champ spatial nécessaire pour appréhender la scène, confirmant l'hypothèse d'un traitement différent pour les informations périphériques et les informations fovéales. Dans un contexte clinique, certains patients souffrant d'agnosie sont dans l'impossibilité d'identifier les objets d'une scène tandis qu'ils peuvent être capables de classer les scènes, par une activation de la zone para-hypocampale.

2.3.6 Un cas particulier: dissociation entre la perception de l'objet et de la scène

Nous nous intéressons à « l'effet de foule » car il illustre tout d'abord le fait que les informations sur les objets de scènes sont intégrés à différents niveaux, mais aussi qu'il existe des interactions entre objets de scènes limitant l'analyse locale au profit d'une analyse globale. Cet effet se manifeste lorsqu'un élément visuel exposé individuellement est correctement perçu alors que ce même élément exposé dans un ensemble devient impossible à traiter. La perception d'un ensemble peut donc être différente de la perception de chaque partie de l'ensemble. En effet, des interférences perceptives entre éléments spatialement proches produisent un effet de masquage dégradant la représentation individuelle, alors que le signal perceptif lui-même n'a pas été dégradé.

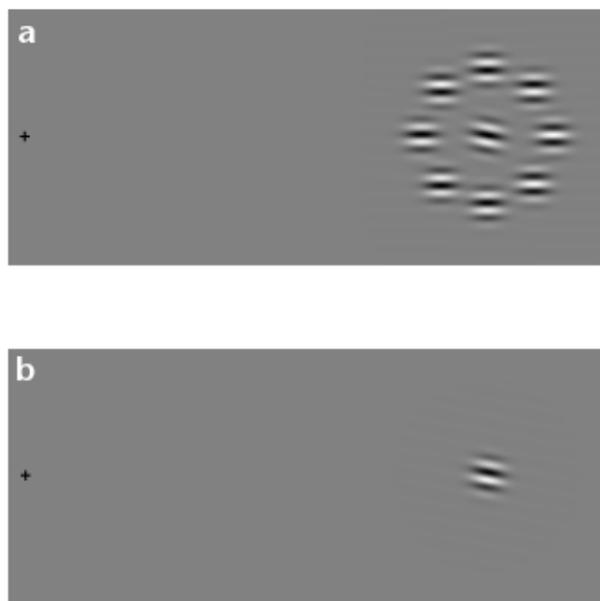


Figure 9 Illustration du « crowding effect » : la fixation de la croix dans la figure a ne permet pas de définir l'orientation du patch central, tandis que cela est possible pour la figure b / [Cavanagh01]

Sur l'exemple précédent (Figure 9, voir [Cavanagh01]), lorsque le point noir de la figure 'a' est fixé, il n'est pas possible de définir l'orientation spatiale du patch de Gabor⁷ central. Cependant, lorsque le point noir de la figure 'b' est fixé, cette orientation devient perceptible. Ainsi, alors même que le signal visuel est identique en terme de contraste, intensité, forme, taille, et excentricité fovéale, l'information est dans un cas disponible et dans l'autre ne l'est pas. Cette perte d'information, jusqu'à présent considérée comme issue d'un filtrage sélectif de bas niveau, est désormais attribuée à un processus positionné dans la chaîne de traitement de l'information visuelle de telle sorte que l'on n'y ait pas accès consciemment. Parkes et al. [Parkes01] montrent que même si les informations portées par chaque élément n'arrivent pas à atteindre un niveau suffisant pour en permettre une description individuelle, ces informations vont néanmoins atteindre des niveaux de traitement suffisamment élevé pour contribuer à l'identification globale. Dans l'exemple précédent, s'il n'est pas possible de définir l'orientation du patch central même s'il est perpendiculaire à l'orientation des

⁷ Du nom du physicien anglais d'origine hongroise Dennis Gabor, un filtre de Gabor est un filtre linéaire défini par le produit entre un filtre Gaussien et une sinusoïdale orientée. Il permet de traiter une petite partie du signal et non sa totalité – source : wikipedia

patches périphériques, il est possible de définir une orientation spatiale globale. Or, cette orientation globale est évaluée par les participants comme valant la moitié de l'angle formé entre les deux valeurs angulaires, du patch central et du patch périphérique. Les informations locales de chaque patch sont donc bien intégrées mais sans qu'il soit possible de les discriminer. Si cet effet ne s'applique pas directement aux types de scènes que l'on exploite, il illustre deux points importants pour l'analyse des processus perceptifs : une information perceptive peut être intégrée sans être identifiée, et le traitement d'une scène ne fournit pas nécessairement les mêmes résultats que le traitement des éléments qui la compose ; ainsi une information visuelle peut participer à l'évaluation globale sans que l'on puisse identifier la valeur de cette information. Cet effet peut aussi s'appliquer à la perception des textures, qui sont typiquement issues d'interactions perceptives (sans identifications) individuelles. L'effet de foule permet ainsi de discriminer très rapidement les textures, qui sont des indices très utiles à la reconnaissance et à la catégorisation des entités visuelles. La perception d'un ensemble plutôt que d'une somme d'éléments, par dégradation des propriétés individuelles, permet d'identifier une texture avec un coût de calcul très faible, et de simplifier l'information visuelle sans qu'il n'y ait de pertes dommageables.

✓ *synthèse*

Ce chapitre a permis de voir que les scènes visuelles peuvent être de différentes natures. On choisit ainsi dans le cadre de notre travail de travailler sur des scènes dépeintes et non des scènes réelles car il s'agit de situations moins limitatives en terme de répétabilité des conditions expérimentales. On dégage ensuite que les scènes dépeintes possèdent plusieurs propriétés, quantifiables selon les axes abstrait/réaliste, naturel/artificiel, et simple/complex. Ces différentes dimensions influenceront sur les processus perceptifs activés. De nombreux niveaux d'analyse de la scène visuelle semblent exister, tant en terme holistique (le degré d'abstraction des informations exploitées) qu'en terme analytique (décomposition de la scène en sous-entités). De nombreux arguments font considérer que le traitement perceptif et cognitif d'un objet est différent de celui d'un ensemble d'objets. Il est donc nécessaire dans des études portant sur des objets manufacturés de prendre en compte le fait, par exemple pour une scène contenant un habitacle de véhicule, que la consigne peut activer un traitement à un niveau global (« entité habitacle ») ou à un niveau local (« planche de bord+volant+sièges »). Enfin, on a pu voir dans la dernière section que la densité informationnelle entourant un stimulus limite la capacité au traitement individuel. Ainsi, on doit prendre en compte le fait que l'évaluation d'une sous-partie d'une scène peut différer selon que cet élément soit présenté seule ou dans son contexte.

La partie suivante va s'intéresser de plus près aux processus perceptifs et cognitifs impliqués dans l'identification des scènes. Afin de comprendre les différents mécanismes permettant d'analyser le contenu d'une scène pour un traitement évaluatif ultérieur, nous nous intéressons aux processus mnésiques et attentionnels liés à l'identification des scènes, puis aux interactions entre reconnaissance des objets et reconnaissance des scènes.

Chapitre 3

Mécanismes cognitifs de l'évaluation visuelle

3.1 Mémoires visuelles et Processus d'identification des scènes

3.1.1 Introduction

La problématique des processus d'identification des scènes visuelles est largement abordée par les paradigmes d'affichage rapide. Ce type de dispositifs consiste à exposer une image sur une durée permettant de n'effectuer qu'une seule fixation du regard, soit environ 150ms. La consigne est soit de détecter soit d'identifier une scène ou un élément de scène et l'exercice est répété de nombreuses fois (« Rapid serial visual display – RSVP », [Williams98], [Potter76], [Intraub81], [Thorpe 97]). Les principales observations sont les suivantes : il serait possible en une simple fixation d'extraire l'identité d'une scène réelle (temps d'acquisition de l'information faible) [Intraub81] ; il serait possible de dire si l'image observée a été vue précédemment (bonne rétention de l'information globale) [Potter76] ; la détection d'un objet dans l'image serait corrélée à sa cohérence dans son contexte et l'identification d'un objet est dépendante de l'identification de la scène (deux niveaux d'informations imbriqués [Biederman83], mais ce point doit être discuté plus loin); la déstructuration de la scène sans décomposer les objets principaux altérerait son identification [Biederman82]. Ainsi, les processus d'identification impliqueraient différents niveaux de traitement des informations présentes dans la scène (organisation spatiale, critères diagnostic, ...) et divers types d'encodage des informations en mémoire (configuration spatiale, sémantique,...). Dans cette partie nous présentons dans un premier temps les processus mnésiques impliqués dans la perception visuelle, puis nous nous intéressons aux processus d'identification des scènes, et enfin à la relation entre la scène et les objets de scène dans la reconnaissance.

3.1.2 Processus de mémorisation des scènes et objets de scène

Dans la vie de tous les jours, il est utile d'avoir accès à plus d'informations visuelles que celles fournies par la position courante du regard à un instant donné, et de se rappeler les détails d'un objet observé pour une tâche ultérieure. Il a été suggéré que le système visuel construit une copie conforme de la scène par l'assemblage des différentes informations visuelles traitées par chaque fixation ([Rayner78] [Marr82]). Cela permet d'expliquer que nous percevons un monde visuel stable malgré les mouvements de la tête et du regard. Cependant, de nombreux éléments ont été avancés pour infirmer cette hypothèse de construction trans-saccadique d'une image mnésique fidèle à la réalité. Tout d'abord, les participants ne sont pas capables de détecter le déplacement d'un objet dans la scène si le déplacement se produit pendant une saccade [Bridgeman75]. Puisque le flux visuel est inhibé pendant la saccade, cela signifie que l'on ne perçoit pas de différences avant et après la saccade, et donc que la représentation cognitive de la scène n'est pas de haute fidélité. Ensuite, Rayner et Pollatsek [Rayner83] observent que les participants ne peuvent reconstituer des

formes composées de points si une partie des points s'affiche pendant une fixation et la deuxième partie pendant une autre fixation. Il n'y a donc pas eu de stabilité mnésique entre les deux fixations. Ensuite, les études de détection du changement montrent que les participants détectent difficilement les changements de couleurs ou de disparition d'objets si le changement coïncide avec une saccade ([Grimes96], [McConkie96]), ou avec un clignement de paupière [O'Regan00], ou une interruption par masquage ([Rensink00b]). Il paraît donc peu vraisemblable que le système visuel construise une représentation interne fidèle de la scène. Malgré tout, de nombreuses informations visuelles sont bien encodées en mémoire ([Henderson97], [Pollatsek92], [O'Regan83]). Les propriétés globales de la scène visuelle telles que le sens (« le résumé » - [Biederman82]) ou le schéma (« la composition spatiale » - [Simons96], [Gibson79]) sont rapidement et stablement encodées en mémoire. Les propriétés des objets peuvent aussi être retenues, telles que leur identité, leur forme, leur couleur ([Henderson99], [Hollingworth02], [Tatler03]). C'est pourquoi de nombreux auteurs ont porté leurs efforts sur la compréhension des différents types de mémoires visuelles interagissant avec l'attention et l'orientation du regard.

Dans un premier temps, nous distinguerons trois types de mémoires visuelles : celles encodant les connaissances liées à la tâche, celles encodant les informations liées à la pré-connaissance de la scène, et celles rapportées aux connaissances immédiates de la scène, c'est-à-dire sa représentation. Puis nous présenterons quelques unes des théories portant sur l'encodage en mémoire des objets.

3.1.2.1 Mémoire de tâche

Ce type de connaissances consiste à encoder en mémoire des procédures d'exploration visuelle pour une tâche donnée. Cette mémoire intègre des règles contrôlant le regard, autrement dit des stratégies d'exploration optimisées (ex : regarder dans son rétroviseur pendant une tâche de conduite), et permet d'anticiper la position du regard pour les demandes cognitives de résolution de la tâche. Ainsi, selon que l'on demande au participant de mémoriser une scène ou de rechercher un objet, la distribution des fixations va évoluer avec les répétitions, optimisant la tâche.

3.1.2.2 Mémoire visuelle photographique de scène

✓ *La mémoire transitoire*

Henderson et Ferrera [Henderson04a] postulent l'existence d'une mémoire visuelle transitoire de scène, stockant des informations spécifiques de la scène en cours de traitement par l'attention. Les objets seraient ainsi fixés selon les besoins du système cognitif. Par exemple, je sais que mon téléphone mobile est actuellement posé à gauche de mon clavier, et mon regard peut s'y poser directement si la sonnerie retentit. Cette mémoire sert aussi à réorienter le regard vers les zones de la scène qui sont informatives ([Buswell35], [Yarbus67]) ainsi qu'à détecter les nouveaux objets ou des objets disparus de la scène [Brockmole05].

✓ *La Mémoire stable*

Il semble de plus qu'il existe une mémoire visuelle photographique de scène à long terme. Ce type de mémoire stocke les informations redondantes contenues dans une scène observée plusieurs fois. Par exemple mes clés sont « toujours ou presque » placés à proximité de la porte ; je vais donc diriger le regard dans cette direction pour les retrouver. Le processus de « contextual cueing » ('ciblage contextuel') se base sur ce type de mémoire. Le ciblage contextuel est un apprentissage perceptif qui apparaît très rapidement [Chun98] et qui influence directement les mouvements des yeux [Peterson01]. En effet, ce type de processus basé sur la mémoire à long terme optimise la recherche d'informations perceptives pour détecter une cible. Par exemple, pour une tâche de recherche d'un objet dans une image, si pour plusieurs essais, quelle que soit la cible, celle-ci est sur une table en bois, alors dès que la table en bois sera localisée, les yeux s'y poseront. L'environnement perceptif est alors utilisé en tant qu'indice de prédiction de cible. Ce type de processus est tout aussi efficace pour

les scènes abstraites que pour les scènes réelles. Brokmole et Henderson [Brockmole06] montrent que les représentations acquises par l'apprentissage contextuel facilitent les mouvements oculaires dans les tâches de recherche de cible.

3.1.2.3 Mémoire du schéma et résumé de scène

✓ *Exploitation des invariants*

La mémoire du schéma et du résumé de scène comporte les informations sémantiques et spatiales à propos d'une catégorie spécifique de scène [Biederman82]. La mémoire schématique liste les objets vraisemblables d'être trouvés dans une catégorie de scène (ex : un volant dans une voiture), les régularités spatiales et les connaissances des lois invariantes (i.e. la gravité implique qu'un objet peut être posé au sol mais pas au plafond). La mémoire du résumé de scène est de type lexical, permettant de catégoriser la scène. Ces sources d'informations sont utilisées pour optimiser l'exploration visuelle en associant les propriétés de la scène (sa composition, sa catégorie,...) à sa représentation. L'identité de la scène rapidement acquise, les fixations suivantes peuvent être limitées aux zones qui sont susceptibles de contenir le plus d'informations pertinentes pour la consigne.

✓ *Interférences mnémo-perceptive*

Les travaux d'Intraub et collaborateurs ([Intraub93], [Intraub98]) montrent que les schémas de scène jouent un rôle fondamental dans la mémorisation des scènes visuelles. Le phénomène d'extension des frontières en est une illustration. Dans une première phase expérimentale, une photo de scène réelle (pièce d'une maison) est présentée à tous les participants. Dans une deuxième phase, la photo est réexposée, mais accompagnée de plusieurs autres photos de la même scène, prises du même point de vue mais sous différentes focales (i.e. la scène apparaît comme plus ou moins « zoomée »). Les participants doivent alors désigner l'image vue dans la première phase. De manière significative, la majorité des participants considèrent avoir vu la plus large photo de scène en premier, ce qui est erroné. Tout se passe comme si l'activation du schéma de scène de la scène en cours interférait avec la mémoire visuelle à court terme. Les participants se remémorent ainsi des objets qu'ils n'ont pas vus au préalable mais plausibles voire évidents dans la scène. Remarquons que ce phénomène est autant présent pour une durée d'exposition de 15 secondes que de 300ms, indiquant que l'activation du schéma de scène est spécialement rapide [Intraub89].

3.1.2.4 Représentation mnésique des objets de scène

Plusieurs théories présentent différents modèles de la mémoire des objets. La théorie la plus extrême est celle de O'Regan [O'Regan92], proposant que le monde soit une mémoire externe, et considérant qu'il n'y a pas de mémoire visuelle des objets puisque l'information est à portée de regard. La théorie de la cohérence, développée par Rensink ([Rensink00a], [Rensink00b]), propose que seuls les objets cohérents peuvent être représentés en mémoire, mais seulement tant que l'attention s'y attache, allant dans le sens d'une proposition de Wolfe [Wolfe96], pour lequel la description des objets en mémoire est transitoire. Cependant, ce modèle est pris en défaut car il a été noté, notamment par Henderson [Henderson99] que de nombreux items de la scène peuvent être aisément remémorés et décrits plusieurs heures après exposition. Une seconde proposition de modèle considérant la représentation des objets en mémoire comme transitoire est celle de la théorie des fichiers d'objets. Cette théorie (« object file theory » [Irwin92], [Irwin96]) postule qu'il existe en mémoire des fichiers, chacun contenant la description des propriétés d'un objet. Le fichier serait maintenu en mémoire visuelle à court terme même après que l'attention ait quitté l'objet, mais la capacité de cette mémoire serait limitée à seulement trois ou quatre objets récents. Ainsi, l'encodage de l'information serait transitoire mais présent pendant plusieurs fixations selon l'évolution de l'attention, jusqu'à ce qu'un nouvel objet efface l'ancien fichier. Irwin [Irwin02] montre ainsi que la capacité à se remémorer des objets diminue lorsque le nombre de fixations séparant la fixation de

l'item et la remémoration augmente, avec de très faibles performances après trois fixations. Si cette théorie semble pertinente pour des scènes comportant un très grand nombre d'éléments ou des éléments abstraits ou désordonnés, cela n'explique pas comment on peut se souvenir en détail par exemple du contenu de notre cuisine. Ainsi, en contraste à ces observations de rétention d'information transitoire, Henderson et collaborateurs proposent que la description des objets en mémoire puisse être non seulement stable mais suffisamment puissante pour encoder de nombreux objets ([Henderson99], [Hollingworth02], [Hollingworth01b], [Hollingworth06]). En utilisant un paradigme de cécité au changement sur des scènes réelles d'intérieur, ces auteurs ne trouvent pas de preuves de l'existence d'une instabilité temporelle de la mémoire et démontrent que la mémoire des objets visuels est plus performante que ce qui était considéré auparavant. Notamment, les participants détectent aisément des changements d'objets qui interviennent pendant l'exécution d'une saccade éloignée de l'objet disparu. Ce résultat s'oppose à l'hypothèse de mémoire transitoire car, immédiatement avant la saccade, l'attention est censée être dirigée vers la cible de la saccade ([Deubel96], [Hoffman95], [Shepherd86]). Ainsi, au moment du changement, l'attention serait déjà éloignée de l'objet qui va disparaître et donc la description de l'objet ne devrait plus être disponible en mémoire. Si le changement est perçu, c'est que la description de l'objet en mémoire est toujours active, réfutant l'idée qu'une évolution de l'attention efface les objets précédemment encodés en mémoire.

Ces différents modèles pourraient être en fait tous valides, mais chacun pour un certain paradigme. La description transitoire des objets en mémoire serait avantageuse dans les situations où l'environnement visuel évolue alors que la description stable réduit les besoins de ré-extraction des informations visuelles dans un monde statique. Pour des objets abstraits la description transitoire semble adaptée car la représentation de l'objet est coûteuse (pas de gain dû à la pré-connaissance, encodage de nombreuses informations...) alors que pour des scènes réelles cohérentes, l'encodage pourrait consister en une liste des objets présents. On pourrait comparer l'encodage transitoire à un fichier d'image de type bitmap, où chaque pixel de l'image encode une information, permettant de stocker la globalité des propriétés de l'élément mémorisé, mais sans optimisation du stockage. Au contraire, l'encodage à long terme pourrait être un encodage de type vectoriel, où ce ne sont pas les propriétés locales qui sont stockées, mais les lignes principales et les positions des points d'encrage spécifiques ; cela permet un stockage beaucoup plus léger, mais nécessite un niveau d'analyse plus complexe. Dans le cas de la mémoire humaine, c'est la mémoire de scène qui pourrait faciliter cette analyse.

Le traitement des scènes visuelles peut donc se décomposer en au moins trois phases : la prise d'information, qui permet d'extraire et filtrer les informations qui parviendront au système cognitif, l'encodage, qui permet de stocker le flux perceptif en mémoire sous forme de représentation visuelle ou lexical, et la stabilisation de l'information, qui permet de retenir à long terme une partie des informations extraites. Dans le paragraphe suivant, nous nous intéressons aux liens qui unissent l'encodage des informations visuelles dans les différentes mémoires listées avec le comportement oculomoteur.

3.1.2.5 Mémoire et comportement du regard

Les informations concernant les objets dans des scènes réelles complexes semblent s'accumuler en mémoire au cours du temps d'exposition de la scène ([Melcher01], [Tatler03]). Une hypothèse serait que l'accumulation de l'information par la mémoire suivrait l'échantillonnage discontinu imposé par les saccades. Loftus [Loftus81] montre ainsi que les performances de mémoire de l'information visuelle augmentent avec le nombre de fixations lorsque les participants sont autorisés à investiguer l'ensemble de l'image. De même, l'étude de la détection du changement et de l'identification d'objets montre qu'il est en général nécessaire de fixer à proximité ou dessus la région incriminée, afin de percevoir tous les détails ([Hollingworth01a], [Hollingworth01b]). Les fixations seraient donc reliées à

l'encodage en mémoire, tant pour le long terme que le court terme. [Hollingworth02]. Différentes études investiguent en profondeur le rôle des positions des fixations dans la constitution de la mémoire de scène ([Antes81], [Currie00], [DeGræf90], [Hollingworth02], [Irwin02], [Nelson80]). Elles montrent que les positions des fixations influencent l'encodage en mémoire des informations de la scène, mais ne montrent pas que seules ces informations sont encodées [Irwin96]. De plus, la capacité à décrire la scène, verbalement et en détail, augmente avec le nombre de fixations [Loftus72] et la performance de remémoration des objets de scène est lié au nombre de fixations effectuées [Irwin02]. Hollingworth [Hollingworth02] montre dans une tâche de détection du changement sur des scènes réelles que les performances de détection augmentent pour les objets fixés avant le changement, c'est-à-dire que lorsque le temps d'observation d'un objet augmente, la détection du changement est facilitée. S'il est donc clair qu'il existe un lien entre extraction de l'information et position du regard pour une tâche de détection du changement sur une scène réelle, cela ne suggère cependant pas que seuls les objets fixés soient encodés en mémoire.

Tatler et collaborateurs [Tatler05b] testent la capacité à se rappeler différents objets et propriétés d'objets lorsque les participants observent des scènes réelles complexes ou des représentations photo réalistes de ces mêmes scènes. Ils comparent les performances de remémoration des objets en fonction du nombre de fois que l'objet ait été fixé et de la durée séparant la fixation d'un objet à une question, pour les cinq propriétés suivantes : présence de l'objet, forme, couleur, position, distance relative. Afin de comparer l'hypothèse d'un encodage différent selon la nature de la scène, ils effectuent deux expériences, l'une sur scène réelle, l'autre sur photos de ces mêmes scènes réelles.

Dans les deux situations les formes et les distances relatives ont été peu retenues. Pour les scènes réelles, les informations de couleur et de présence d'objets sont acquises en une seule fixation, sans corrélation avec la durée de fixation, tandis que les informations de position s'optimisent avec le nombre et le temps des fixations sur l'objet. De plus, il apparaît que les informations de couleur et de présence sont impactées par la récence de la fixation, la performance diminuant avec le temps, tandis que la position des objets ne semble pas être une information transitoire. Pour les scènes virtuelles, la couleur n'est pas une information transitoire, contrairement aux informations de présence et de position. En termes de nombre de fixations, couleur et présence semblent encodées en une seule fixation, voire sans fixation directe, tandis que la position nécessite plusieurs passages pour être retenues. Des différences entre situations réelle *versus* situation virtuelle apparaissent : dans la situation réelle, les informations de couleur sont transitoires mais les positions sont stables. A l'inverse, en situation virtuelle, les informations de couleurs sont stables mais celles des positions sont transitoires. Ces observations remettent en cause le modèle de Rensink [Rensink00b] car d'une part certaines informations restent disponibles après quelques fixations et d'autre part car les informations de position se construisent au fur et à mesure des fixations. Ces résultats sont cohérents avec le modèle d'Irwin et Andrews [Irwin96] car la présence et l'identité des objets sont encodées sans fixations directes ou en une seule fixation tandis que l'information de position est transitoire et nécessite plusieurs fixations. Les fichiers d'objets pourraient ainsi encoder l'identité mais pas la position des objets. Cependant, le modèle d'Irwin et Andrews propose qu'il n'y ait pas d'informations encodées sur plusieurs fixations alors que cela est observé par Tatler. De plus, certaines propriétés peuvent être stockées à long terme, confirmant la proposition d'Henderson de l'existence d'une représentation de la scène beaucoup plus riche et stable que l'encodage de 3 ou 4 éléments.

Il apparaît des résultats de Tatler que toutes les propriétés des objets ne sont pas traitées de la même manière selon la nature de la scène. La présence et la couleur peuvent être encodées depuis les informations périphériques, et ne s'accumulent pas au fur et à mesure de l'observation, contrairement à la position qui nécessite d'être extraite de manière fovéale et plusieurs fois fixée. Dans le monde réel, les propriétés des objets telles que l'identité et la couleur changent peu, ce qui implique qu'un retraitement est inutile. Cependant les objets du monde réel étant mobiles, la position

doit être réintégrée régulièrement au cours de l'action, donc n'est que transitoire. Par leur nature, le monde réel et virtuel se différencie. En effet, pour une scène réelle, le champ de vision est plus large, les informations de profondeur sont présentes, la dynamique de luminance est supérieure, et les objets sont mobiles. Ainsi, les moyennes des durées de fixations sont supérieures pour les scènes réelles, et le nombre de fixations est plus faible, malgré un temps d'exposition équivalent. La description des objets encodée en mémoire différencierait donc selon la situation écologique.

Après s'être intéressé aux processus mnésiques, il est nécessaire pour notre étude d'aborder la problématique de l'identification des scènes visuelle et de leur contenu.

3.1.3 Processus d'identification des scènes

3.1.3.1 Schéma et résumé de scène

L'identification de la scène consiste en un processus qui requiert peu d'effort attentionnel [Li02], rapide car possible en une seule fixation [Biederman82], sollicitant fortement la mémoire, et pouvant se baser sur les statistiques « signales » globales de l'image [Torralba03b]. Ces statistiques « signales » seraient prédictives de l'identité de la scène et du contenu sémantique. Au moins deux types de connaissances sont activés pour l'identification de la scène : le résumé de scène (« gist ») et son arrangement spatial (schéma, « layout ») [Henderson04a]. Le résumé de scène fait référence à la signification de la scène et à sa catégorie conceptuelle de base (ex : un habitacle de véhicule). Il est activé par le traitement des informations perceptives « signales », de la détection d'objets clés, et de la sémantique de l'ensemble [Oliva04]. Ainsi des catégories telles que « plage », « rue », « forêt » [Tversky83] peuvent être reconnues en seulement 45 à 135ms [Schyns94]. Le résumé représente ce qu'est la scène et peut être décrit en une simple phrase : passants dans une rue, parking, etc. Il s'agit d'une caractérisation par des descripteurs de niveau global, à l'inverse d'une description détaillée de chaque objet de la scène. Le résumé est donc une approche holistique de la scène, et non un traitement local analytique. L'arrangement spatial (schéma) de scène est un ensemble de connaissances liées à la disposition de la scène, de ses différents plans, de ses objets, ainsi que leurs tailles relatives. Les distances relatives des objets et leurs formes y sont aussi encodées. Lorsque cette connaissance est activée, l'attention optimise le traitement informationnel sur les zones prédites comme pertinentes. Il s'agirait d'un processus de traitement parallèle produisant un guidage attentionnel rapide ([Treisman98], [Biederman81], [Hess90]) et permettant la détection d'objets prédits [Germeys01].

3.1.3.2 Détecter et identifier, deux opérations distinctes

Le délai nécessaire à l'identification d'une scène réelle complexe est très court. Biederman [Biederman98] montre qu'en une seule fixation (de 133 à 300ms), la catégorie de la scène est acquise, son résumé lexical et son schéma étant ainsi disponibles à l'attention. Une description verbale du type « un bateau près d'un ponton » est possible dans ces délais [Potter75]. Des images affichées pendant des durées très brèves, inférieures à 120ms [Potter76], peuvent aisément être reconnues *a posteriori* dans une tâche de reconnaissance si le délai consécutif à l'exposition de l'image est suffisamment long pour permettre une consolidation en mémoire des informations. Les travaux de Thorpe et collaborateurs [Thorpe96] montrent quant à eux que 20ms d'exposition peuvent être suffisantes pour détecter la présence d'un animal dans la scène, même si la décision est prise au bout de 150ms et que l'identification de l'animal n'est pas possible. Il est important de faire la distinction entre détection et identification. Dans le cas des expositions ultra courtes, les participants sont capables de détecter une cible, mais pas nécessairement de l'identifier. Cela signifie que la présence ou non d'un animal dans la scène sera correctement analysée, sans qu'il soit possible d'explicitement de quel animal il s'agit. Une conséquence du délai nécessaire à l'identification implique que si de nouvelles images sont exposées dans les 150ms suivant la disparition de la première image

[Rousselet02], ces informations d'identification sont court-circuitées par les données d'identification du nouveau stimulus. Friedman propose donc que dans le cas de la catégorisation rapide, un « objet diagnostique » déclenche l'identification de la scène [Friedman79]; par exemple un ordinateur pour un reconnaître un bureau.

3.1.3.3 Processus d'identification des scènes : traitement visuel diagnostique ou holistique ?

La reconnaissance des scènes peut être produite par plusieurs approches allant de la détection d'objets diagnostiques à la perception d'une signature spectrale globale. Il est ainsi pertinent d'observer comment se répartissent les processus d'identification basés sur des indices locaux avec les processus basés sur des informations globales.

Afin d'observer si la reconnaissance des scènes naturelles s'effectue par un diagnostic local ou une signature globale, Vogel et Schwaninger [Vogel06] recueillent 250 images de scènes naturelles et définissent 5 catégories : côtes, rivières/lacs, forêts, plaines, montagnes. Dans une première expérience, les participants sont exposés pendant 4 secondes à chaque image et doivent ensuite les classer dans l'une des catégories. Les résultats montrent un taux moyen de réussite de 89,7%. Ils renouvellent l'expérience mais chaque image est découpée en 100 parties de tailles égales et réassemblée aléatoirement. Ainsi, des éléments sémantiques tels qu'eau, sable, rochers, sont toujours présents en terme texturaux mais la signification globale de la scène est incohérente. Le taux de réussite de la reconnaissance diminue à 72,7% (avec un écart type dépendant de la catégorie). Enfin, le même test est effectué mais chaque image naturelle d'origine est traitée par un filtre passe-bas. Le taux de réussite de reconnaissance est alors de 71,6% (écart type ici aussi dépendant de la catégorie). Les auteurs notent alors que lorsque le taux de reconnaissance est faible pour la version de scène déstructurée, il est alors élevé pour l'image filtrée. La reconnaissance des scènes naturelles ne semble donc pas uniquement liée aux informations locales ni uniquement aux informations globales, puisque les taux de réussites de catégorisation sont identiques pour les situations de dégradation locale et globale, mais que ces taux sont inférieurs au taux de réussite des images non modifiées. De plus, la nature de la scène influe sur le type de processus sollicités, car certaines scènes sont mieux reconnues avec les informations locales et d'autres avec les informations globales. Il n'est donc pas possible d'affirmer que le mode diagnostic est celui qui permet la catégorisation visuelle, ni de dire qu'il n'a pas d'influence. Il semble que les effets respectifs de ces deux facteurs soient pondérés selon le type de sémantique de scène. Notons que lorsque l'identification se fait sur la base d'informations de basse fréquence, on considère qu'il s'agit d'une reconnaissance par « signature ».

L'hypothèse de l'approche « diagnostic » postule qu'un ou plusieurs objets soient rapidement et parallèlement identifiés, inférant le résumé et schéma de scène ([Friedman79], [DeGræf90], [Biederman82]). Cette approche se base sur l'idée que la reconnaissance visuelle repose sur les mêmes processus que ceux permettant la reconnaissance des scènes. Mandler et Parker [Mandler76] proposent ainsi que trois types d'informations sont analysées : inventaire d'objets, informations descriptives de l'apparence et des détails de certains objets, relation spatiale entre les objets. Wolf [Wolfe98a] ajoute à ces informations d'objets, des informations portant sur les relations sémantiques entre objets, l'organisation spatiale et l'impression générale des caractéristiques perceptives et texturales. Une limite se pose cependant car le résumé de scène peut être extrait d'une image dégradée ou dont les détails ont été filtrés ([Schyns94], [Mannan95]), ce qui signifie que ce ne sont pas des détails de diagnostic qui ont aidés à l'identification. Notons que le modèle « functional isolation model » [Hollingworth99] propose que l'identification des objets de scène soit indépendante des attentes issues des connaissances de scène. Cet auteur n'observe ainsi aucunes relations systématiques entre les performances de reconnaissance de scène et reconnaissance des objets.

La reconnaissance rapide des scènes ne nécessite pas nécessairement de traiter les objets de scène. Ainsi, lors des premières phases de l'exposition, les mécanismes de reconnaissance des scènes réelles ne semblent pas être les mêmes que ceux de la reconnaissance des objets ([McCotter05], [Fei-Fei04]). En 1983, une étude de Loftus et collaborateurs [Loftus83] portant sur le traitement des informations holistiques *versus* locales lors de tâches de reconnaissances d'images laisse apparaître que les participants sont plus performants lorsque leurs réponses s'appuient sur des cibles holistiques. Une seule fixation du regard et non pas un scan entier de l'image est alors suffisant pour reconnaître une image. Une image de scène serait initialement traitée en tant qu'entité tandis que les informations locales à propos des objets ne seraient traitées que plus tardivement, lorsque l'évaluation du contexte de scène permettrait d'orienter le déploiement de l'attention sur des zones spécifiques. Ainsi, Oliva et collaborateurs [Oliva97] construisent des images hybrides en combinant informations de basses fréquences spatiales d'une première photographie avec les informations de hautes fréquences d'une autre seconde photographie. En 50ms, les participants identifient correctement la scène ayant fourni les basses fréquences. L'identité est donc extraite ici des basses fréquences et ce sont ces informations qui sont reconnues en premier (avant les hautes fréquences, donc les objets).

3.1.3.4 Exploitation computationnelle des indices d'identification des scènes

Le modèle d'enveloppe spatiale fournit l'une des explications potentielles des mécanismes de reconnaissance globale (holistique) des scènes. L'enveloppe spatiale est issue des propriétés fréquentielles des scènes telles que la rugosité texturale, l'étendue, ou l'ouverture. Il ne s'agit donc pas d'une approche analytique basée sur les primitives visuelles (lignes puis formes puis objets de scènes), mais sur une signature globale à laquelle on attribue une connotation sémantique. La modélisation computationnelle de cette approche permet de catégoriser un grand nombre de scènes réelles selon leur contenu sémantique [Oliva01], [Oliva03], [Greene06], [Oliva06]. Dans une étude de 1999 (Oliva, Torralba, Guérin-Dugué et Héroult [Oliva99]), 700 photographies de scènes réelles complexes sont positionnées dans un espace à 3 dimensions : artificiel/naturel, ouvert/fermé, étendu/clôturé. Le calcul de la réponse du spectre d'image à des filtres de Gabor permet de classer dans cet espace 90% des images. Guérin-Dugué et collaborateurs [Guerin00] utilisent les orientations des spectres des images pour une catégorisation en quatre catégories : scènes d'intérieur, scène urbaines, scènes ouvertes, paysages. 470 images en niveau de gris sont analysées en calculant pour chaque point de l'image l'orientation dominante et l'énergie issues de filtres orientés et ce pour différentes fréquences spatiales ; le tout forme un histogramme par fréquence spatiale. Une matrice de distance permet ensuite de retrouver les catégories pour 82 à 90% des essais. Enfin, Torralba & Oliva [Torralba99] proposent de positionner des photographies de scènes naturelles dans un espace sémantique en fonction du degré d'ouverture de la scène, qui est corrélé au niveau d'énergie calculée par des filtres linéaires discriminants. La première dimension de cet espace sémantique est le degré de naturalité de la scène (représentant des objets manufacturés par opposition à celles représentant des paysages) ; une scène représentant une cabane dans une forêt est ainsi positionnée entre les deux extrémités. La deuxième dimension est celle du degré d'ouverture, estimé selon la quantité de lignes verticales et horizontales. L'axe va du vertical (building) à l'horizontal (autoroutes). Le filtrage par filtre de Gabor permet de distinguer ces deux dimensions et de classer automatiquement les images de scènes naturelles selon ces caractéristiques. Ainsi, le degré de naturalité de la scène peut s'identifier par des informations « signal » globales ; or il s'agit d'un classement sémantique. On peut donc envisager que le système visuel identifie la sémantique globale de la scène par un traitement rapide sur le signal, sans identification locale. Cependant l'approche opposée est aussi vérifiée, c'est-à-dire que des indices locaux de nature « signal » peuvent permettre une classification sémantique. Ainsi, dans une approche à mi chemin entre objet diagnostique et spectre visuel, Renninger ([Renninger03], [Renninger (04-07)]) demande à des participants de classer des scènes naturelles après 37 à 69ms d'exposition selon leurs propriétés sémantiques. Le taux de réussite étant

significativement au dessus de la chance, cette étude confirme l'existence de processus diagnostic efficaces. Elle propose alors qu'une simple analyse de texture des images peut permettre d'identifier la scène. Un algorithme d'apprentissage computationnel basé sur les théories de l'information établit un histogramme de fréquences d'unités texturales (« textons ») par catégories sémantiques, en apprenant les caractéristiques de textures de 10 scènes de bases. En comparant la scène exposée avec la base de données des textons, le modèle identifie la sémantique de base de la scène naturelle dans des délais similaires aux performances humaines.

3.1.3.5 Interférences cognitives dans l'identification des scènes

✓ *Effet des incohérences sur l'identification des scènes*

Biederman [Biederman72] applique un paradigme de décomposition aléatoire de l'image, c'est-à-dire que l'image est décomposée en zones géométriques (6 parties), et les positions de ces zones géométriques sont interverties aléatoirement. Les objets sont positionnés dans l'image de telle sorte qu'ils ne soient pas coupés par le quadrillage. Un objet cible est défini, et la zone le contenant est la seule qui ne soit pas déplacée. Il apparaît que le temps de détection de la cible est plus élevé pour les scènes décomposées (« jumbled ») que pour les scènes originales, alors même que la sémantique des objets n'a pas été impactée. Il serait donc plus difficile de reconnaître et de localiser des objets dans une scène désassemblée que dans une scène cohérente. Cet effet d'incohérence s'observe aussi lorsqu'une information contenue par la scène est inattendue ; l'attention et l'identification s'en trouvent alors impactées. Biederman [Biederman82] analyse ces effets. Ainsi, dans un schéma de scène, si le support (lois de la nature, ex : gravité), l'interposition (superposition, inclusion d'objets), la position (c'est-à-dire qu'il est vraisemblable de trouver cet objet, mais pas dans ce lieu précis de l'image), et la taille (c'est-à-dire que l'objet est trop petit ou trop grand) ne correspondent pas aux pré connaissances, alors ces informations locales sont considérées comme « incohérentes » : elles perturbent la reconnaissance et modifient l'orientation de l'attention. De même, s'il existe dans la scène un objet n'étant pas sémantiquement adéquat au contexte, ou si sa probabilité de présence est faible, alors une incohérence va être détectée. Davenport [Davenport04] montre que les objets cohérents dans la scène soient identifiés plus aisément et plus rapidement que les objets incohérents. Cependant, en terme oculomoteur, Loftus et Mackworth [Loftus78] présentent des dessins comportant des objets incohérents, et constatent qu'ils sont fixés par le regard avant et plus longuement que les objets cohérents (voir aussi Friedman [Friedman79]). Il y aurait donc dissociation entre l'identification et les fixations. Le regard, par les premières fixations, s'orienterait vers les éléments incohérents. Underwood truque des photos de scènes réelles en plaçant des objets incongrus dans certaines images. Les participantes doivent effectuer une tâche de recherche de cible. Les objets incohérents sont détectés en moyenne en 2.8 fixations (0.97s ; 1.8 saccades) tandis que les objets cohérents sont trouvés en 3.4 fixations (1.11s, 2.4 saccades). Une fois détectés, les objets incohérents sont fixés 1.39s contre 1.07s pour les objets cohérents. Henderson observe que la première et deuxième fixation ainsi que leurs durées totales sont plus longues sur les objets incohérents ou « anormaux » que sur les objets cohérents sémantiquement avec le contexte. Les objets incohérents avec le contexte attireraient l'attention rapidement et engendreraient de longues fixations, mais seraient aussi plus difficiles à identifier et pas nécessairement informatifs pour la tâche. Ainsi, moins un objet correspond au schéma et résumé de scène, plus il est attractif, et moins est rapide son identification.

Les travaux de Bar [Bar04] ou Boyce et Pollatsek [Boyce89] confirment la proposition d'un effet de l'incohérence sur l'identification de la scène tandis qu'Hollingworth et Henderson [Hollingworth98] ainsi que de De Graef [DeGraef98] les infirment. Ils n'observent pas de facilitation de l'identification d'un objet par la cohérence de son contexte. Sur des dessins schématiques issus de photographies de scènes réelles, et pour une tâche de mémorisation, les sujets n'orientent pas spécifiquement leurs yeux en direction d'un objet selon sa nature incohérente ou cohérente avec le contexte. Dans une étude complémentaire de De Graef, les participants sont invités à rechercher une cible dans une scène

de dessins schématiques de photographies le plus rapidement possible, cible définie par un mot avant chaque exposition. Il n'est pas apparu de processus « Pop-Out » (élément attirant fortement l'attention pour des raisons sémantiques dès les débuts de l'exploration de la scène) qui auraient orientés l'attention plus rapidement sur les objets incohérents que sur les objets cohérents. Au contraire, les cibles cohérentes étaient fixées plus tôt que les cibles incohérentes. La reconnaissance de la scène pourrait donc orienter la recherche de la cible sur des lieux probables et non sur l'ensemble de la scène. De même, Mackworth et Morandi [Mackworth67] rapportent que les observateurs tendent à regarder en premier les zones qu'ils considéreront plus tard comme les plus informatives, et pas nécessairement les zones incohérentes. L'incohérence sémantique peut de plus être accompagnée d'une forte distinction visuelle de l'objet par rapport à l'arrière plan, ce qui se vérifie sur les études de Loftus et Mackworth. On doit aussi remarquer que les scènes utilisées par Loftus et Mackworth [Loftus78] étaient moins complexes que les représentations de scènes réelles, facilitant l'extraction des informations de la scène pour ensuite orienter l'attention sur les éléments incohérents. On constate donc des résultats opposés. On peut attribuer ces divergences à différents facteurs : variété du type d'images employées (réalistes/dépeintes, concrètes/abstraites) mais aussi à la consigne (recherche de cible/exploration libre) ou au type d'incohérence (sémantique, contextuelle, violation de lois physiques, etc.). Ce qui semble important dans tous les cas, c'est qu'il peut survenir des processus d'orientation de l'attention distincts selon le degré de cohérence d'un élément de scène mais aussi selon le niveau de cohérence de la globalité de la scène.

✓ *Effets de la sémantique des scènes sur les processus d'identification*

Afin d'observer s'il existe des processus différents de reconnaissance des scènes visuelles selon leur contenu sémantique, Fei Fei et collaborateurs [Fei-Fei07] recueillent des descripteurs lexicaux libres sur un très grand nombre de scènes visuelles. Ils effectuent ensuite la classification de ces descripteurs selon les groupes suivants : objets animés (animal, humain), objets inanimés (naturels, manufacturés), scènes d'extérieur (naturelle, manufacturée), scènes d'intérieur, caractéristiques perceptives (couleur, forme), scène sociale (ex : repas, concert). Puis, chaque image est exposée à des participants avec des durées d'affichage variables. Les images appartenant aux catégories des objets animés (humains par exemple) sont correctement reconnues et classées après 30ms d'exposition, ce qui n'est pas le cas pour les autres catégories. Cette différence de traitement diminue lorsque le temps d'exposition augmente. Une deuxième observation est que pour de brèves durées d'expositions, pour les objets inanimés, les catégories super ordinales (moyen de transport) sont mieux identifiées que les sous catégories plus détaillées (voiture). Ces résultats montrent que les objets et scènes partagent un certain nombre de ressources de traitement et que la perception et l'identification de l'un favorise la reconnaissance de l'autre. Enfin ces auteurs remarquent que les scènes naturelles sont reconnues beaucoup plus rapidement que les scènes d'intérieurs ou que les objets isolément. Ils montrent complémentirement qu'il est possible de distinguer scènes naturelles extérieures et intérieures selon des informations de couleur et de formes ([Fei-Fei05], [Torralba03]).

✓ *Effets de la consigne sur l'identification*

Schyns et oliva [Schyns99] montrent que l'exploitation du spectre fréquentiel dépend de la consigne. Ils observent sur des images hybrides (mélange de hautes fréquences d'une expression faciale avec les basses fréquences d'une seconde expression faciale dans une même image) que selon que l'on demande de catégoriser les visages exposés par le sexe, l'expression (neutre ou expressive) ou le type d'émotion, les fréquences spatiales utilisées ne sont pas les mêmes. Puisque l'image hybride est composée d'une fusion de fréquences spatiales issues de percepts différents, l'image perçue est reconnue et catégorisée différemment. Ainsi dans une tâche de détection d'une émotion (visage expressif/non expressif), si un lot de visages est jugé inexpressif (informations issues des hautes fréquences), ce même lot d'images est évalué comme représentant des expressions joyeuses (basses fréquences) dans une tâche de reconnaissance de l'émotion faciale. L'utilisation des plages de

fréquences dépend donc de la demande d'informations pour la catégorisation. Il y a ainsi une orientation de l'identification d'une scène en fonction du type de catégorisation demandé.

✓ *Effets de la familiarité et de l'expertise sur l'identification*

Il a été montré qu'il existe des ressources neuronales spécifiques pour le traitement visuel des humains ou des parties du corps humain [Kanwisher01], et que ces ressources sont construites par la familiarité. Un visage familier est plus vite reconnu en tant que visage qu'un visage inconnu, car le traitement visuel active immédiatement des processus associés à la catégorie « visages ». La familiarité est ainsi un facteur modulant le coût attentionnel requis pour l'identification des objets [Fei-Fei04] et il semble exister un processus d'optimisation de reconnaissance et de catégorisation des objets familiers. C'est pourquoi les novices et les experts ne détectent pas les mêmes informations. Schyns ([Schyns94], [Schyns97]) propose qu'il s'agisse là aussi d'un processus d'apprentissage de catégories, l'identification de nouvelles propriétés d'un objet produisant la création de nouvelles sous-classes. L'identification visuelle des scènes dépend donc de la familiarité de celles-ci.

3.1.3.6 L'identification perceptive versus linguistique

Sur la base de travaux en linguistique, Rosch [Rosch78] propose que les objets soient inclus dans différents niveaux de catégories : niveau de base, super-ordinaire (abstrait, global) et sub-ordinaire (figuratif, détaillé). Tversky et collaborateurs [Tversky83] proposent la même taxonomie des catégories de scènes, en se basant sur une heuristique de maximisation de l'information linguistique portée par un nom. Les catégories sémantiques s'organisent selon deux dimensions : horizontalement, les différentes classes (ex : véhicule-arbre,..), verticalement, les sous classes (vélo-voiture | chêne-hêtre). Il existerait cependant un niveau privilégié (le niveau « de base »), selon lequel les participants catégorisent le plus rapidement les objets, par rapport à des catégories plus large ou plus restreintes. Mais la catégorisation visuelle suit-elle les mêmes processus et heuristiques que la catégorisation en linguistique ?

Cette modélisation est mise en défaut dans la perception visuelle lorsque l'on mesure la rapidité à laquelle un objet est identifié. Une expérimentation de Macé et collaborateurs [Macé06] montre que lorsque l'on expose des images pouvant contenir des oiseaux typiques (catégorie de base, ex. corbeau), le temps de détection (présence/absence) d'un « animal » est de 65ms inférieur que celui de la détection d'un « oiseau » (391ms vs 450ms). Le taux de bonne réponse n'est cependant pas affecté. Il est donc possible que les catégorisations ordinales de Rosch soient applicables au langage mais pas à la perception visuelle car une catégorisation super ordinaire (animal) est plus rapidement effectuée qu'une catégorisation de base (oiseau) par le système visuel. Macé propose donc que si le niveau privilégié de la catégorisation sémantique est celui de l'objet de base, verbalisé lors d'une tâche de catégorisation (ex : « oiseau »), la catégorisation visuelle se baserait plutôt sur la classe la plus abstraite (super ordinaire, ex : « animal »), indice permettant d'orienter ensuite les traitements visuels détaillés consécutif à l'analyse globale de la scène. La catégorisation visuelle s'appuierait donc sur les premières informations transmises, c'est-à-dire : (1) des informations magnocellulaires, (2) des représentations en noir et blanc, (3) des basses fréquences, (4) les signatures de catégories générales et non d'objets de base, (5) des informations sémantiquement orienté vers l'abstraction plutôt que la concrétisation. Ainsi, la reconnaissance d'une automobile en tant qu'objet super ordinal « moyen de transport » doit être plus rapide que la reconnaissance dans la classe sub-ordinaire « voiture ». Cela conforte la proposition de Schyns et Oliva [Schyns94] que le traitement visuel suit une procédure « coarse to fine », c'est-à-dire du global au local, alors que le traitement linguistique se base sur le temps d'accès mnésique au nom (animal, chien) pour distinguer les différentes catégories hiérarchiques. Notons cependant que dans les tâches de reconnaissance visuelle rapides (Macé), les processus visuels sollicités sont des processus de détection sans verbalisation donc sans effort d'explicitation langagier du choix. Il pourrait y avoir des différences de résultats si une identification

verbale de la nature sémantique de l'objet était demandé (« est-ce un animal ? » vs. « est-ce un oiseau ? »). Certains auteurs [Grill-Spector05] observent cependant que les processus de reconnaissance d'un objet, sont aussi rapides que la détection de leur présence, l'un n'étant donc pas plus coûteux que l'autre.

3.1.3.7 Conclusion

Ces résultats montrent qu'il est important de bien prédéfinir le type d'évaluation qui va être effectuée sur un objet pour identifier les niveaux d'information qui seront traitées, et qu'il existe de nombreux facteurs qui participent à l'identification des scènes et objets.

Après avoir effectué une revue de littérature sur les processus d'identification de scènes, nous abordons dans la partie suivante le concept d'attention, pour dégager les différents processus de guidage du regard en situation d'évaluation d'objets, de nature tant endogène (intrinsèques au système cognitif du participant) qu'exogènes (intrinsèques aux stimuli).

3.2 Attention visuelle et comportement du regard

Dans les sections suivantes, nous nous intéresserons à la nature de l'attention visuelle et à sa corrélation avec le mouvement des yeux et abordons les facteurs qui participent au guidage de l'attention visuelle, tant perceptifs que cognitifs, lors de l'observation active de photographies de scènes réelles.

3.2.1 Propriétés de l'attention visuelle

Afin de pouvoir analyser les données oculométriques pour en extraire des informations sur les informations visuelles traitées, il est nécessaire de cerner le concept d'attention, puis de relier celui-ci aux comportements d'orientation du regard. On va donc chercher à comprendre comment l'attention se déploie sur la scène au cours de l'exposition. Malgré le fait que l'attention puisse être fovéale ou péri fovéale (périphérique), la tendance naturelle est de diriger la zone fovéale vers le lieu de l'attention. Différentes études (voir [Irvin04]) montrent ainsi que les mouvements des yeux sont des indicateurs comportementaux en temps réel des processus attentionnels (exemple en lecture : [Sereno06]). C'est pourquoi dans cette partie nous nous intéressons aux propriétés de l'attention visuelle.

✓ *Caractéristiques*

L'attention peut se décrire sur un continuum dont une extrémité est l'attention focalisée, séquentielle, endogène, et volontaire, et l'autre extrémité est l'attention fractionnée, parallèle, exogène et automatique [Boloix05]. Il s'agit d'une fonction cognitive responsable du filtrage des informations traitées et fournissant à la conscience ce qui est pertinent pour l'observateur. Kowler [Kowler06] propose que l'attention possède plusieurs caractéristiques. La première est d'optimiser la détection d'informations pertinentes dans un ensemble de signaux en diminuant le seuil en deçà duquel l'information n'est pas détectée. La deuxième caractéristique est d'augmenter la perception des contrastes entre les informations, pour fournir une représentation optimisée de la scène. L'attention permettrait de sélectionner un élément spécifique dans la représentation de la scène.

✓ *Les locus perceptifs et attentionnels*

Les mouvements saccadiques du regard sont utilisés pour explorer une scène visuelle car l'acquisition de l'information a lieu principalement pendant les fixations. Cependant, depuis les travaux

de Helmholtz [Helmholtz24] on sait l'observateur capable d'extraire des informations d'une autre zone que celle qui est fixée par le regard. Ce phénomène est nommé attention visible (« overt ») *versus* attention périphérique (« covert »). On considère aujourd'hui que l'attention périphérique n'est pas indépendante de l'attention fovéale, mais plutôt qu'elle complète cette dernière. Ainsi, l'attention pourrait être volontaire, endogène et stable, orientée pour la résolution de la tâche (« overt »), mais aussi automatique, exogène et brève, orientée par les flux perceptifs (« covert »). Posner ([Posner80]) étudia ce phénomène, proposant une métaphore de l'attention en tant que spot lumineux illuminant une région limitée de la représentation mentale de la scène visuelle. Le locus perceptif est défini comme le lieu sur lequel le regard est posé. Le locus cognitif est lui l'objet de l'attention. Lorsque l'attention est fovéale (« overt »), les locus perceptif et cognitif sont confondus. Lorsque l'attention est périphérique (« covert »), le locus attentionnel est périphérique (phase de réflexion).

Les cas les plus courants de dissociation du locus perceptif et cognitif sont les phases de préparation de la saccade. En effet, le mouvement de l'œil est précédé par un saut de l'attention en direction de la cible vers laquelle les yeux vont bouger. Les modèles d'attentions chronologiques comme celui de Henderson [Henderson92] proposent la séquence suivante. Tout d'abord, le stimulus est fovéalisé. Lorsque toute l'information fovéale a été traitée, l'attention est désengagée et allouée en périphérie. La programmation du mouvement de l'œil est alors effectuée. L'attention se repositionne ensuite sur le nouveau lieu à fovéaliser (voir aussi [Posner84]). Cependant cette approche attribue un locus restreint et unique à l'attention visuelle, Kowler [Kowler95] considère que cette proposition n'est pas certaine, en s'appuyant sur des travaux de Findlay [Findlay99] démontrant l'existence de processus de traitement parallèle. Findlay suggère aussi [Findlay01] qu'il n'y a pas de preuves indiquant que l'attention puisse se dissocier de manière à permettre des explorations périphériques séquentielles intra fixations, mais plutôt que l'attention périphérique permet un pré traitement d'informations mises à la disposition de l'attention fovéale.

Van Diepen [VanDiepen98] utilise un paradigme de vision par fenêtre restreinte, floutant l'image hormis la zone sur laquelle la fovéa est orientée. Il apparaît que les performances de reconnaissance de la scène (scène naturelle) sont meilleures lorsque le filtre de flou est passe-bas (c'est-à-dire éliminant les hautes fréquences, donc ne conservant que les zones visuelles homogènes) que passe-haut (c'est-à-dire éliminant les basses fréquences donc ne conservant que les zones avec de nombreux détails, contours, etc.). La vision périphérique fournirait donc plutôt des informations portées par de basses fréquences spatiales, donc plutôt liées à l'organisation spatiale de la scène.

De ces observations nous retenons que le regard est fortement associé à l'attention, sans pour autant pouvoir affirmer qu'une zone fixée soit directement corrélée à un traitement attentionnel. La position des yeux sur la scène révèle donc une probabilité de prise d'informations sans prédire l'intégration ultérieure de ces informations.

✓ *La résolution de l'information visuelle est optimale sur la fovéa mais la prise d'information est globale*

Rappelons que la résolution de la rétine n'est pas homogène sur l'ensemble du champ visuel, du fait de propriétés optiques de la cornée et de la structure neuro-anatomique de la rétine et du cortex visuel. Dans les deux ou trois degrés d'arc centraux du champ central visuel, la qualité du flux visuel est optimale, tout d'abord pour des raisons de densité des récepteurs coniques, parce qu'une grande partie du cortex visuel primaire traite les informations fournies par la fovéa. Ce phénomène nommé *magnification corticale* permet de traiter des flux visuels à haute résolution sur un champ limité. Cependant, Antes [Antes74] montre que les participants qui observent des photographies de scènes réelles effectuent des saccades en moyenne de 6.5 à 8 degrés, suggérant que la pertinence d'une zone est évaluée en vision périphérique (sup. à 3 degrés). De même, Parker [Parker78] demande à

des sujets qui ont pré visualisé des images contenant 6 objets dessinés de les revoir en essayant de détecter les différences. Les sujets sont alors capables de détecter des disparitions d'objets très rapidement : 85% des détections s'effectuant sans fixation sur l'objet ou la zone qui contient l'objet disparu. Ils sont de plus capables de nommer l'objet disparu, confirmant que l'objet proprement dit a été détecté comme manquant et non que seule une modification du schéma de scène a été détectée. Des résultats différents sont cependant observés par Hollingworth [Hollingworth01a] sur un paradigme d'apparition/réapparition de photographies. Plusieurs photographies de scènes naturelles sont effectuées puis retouchées pour supprimer un objet ou le retourner de 90°. L'image originale et l'image retouchée sont ensuite affichées alternativement pendant 600ms chacune, jusqu'à ce que le participant détecte quel objet est modifié. Il apparaît alors que la fovéa doit être à moins de 2 degrés d'angle d'un objet pour détecter son apparition/disparition. Henderson et collaborateurs [Henderson03] montrent sur des photographies de scènes réelles que lorsqu'un changement dans la scène (paradigme de cécité au changement) est détecté, il n'est pas possible de nommer l'objet disparu si le regard est à plus de 4 degrés de l'objet en question, l'identité de l'objet n'ayant donc pas été consolidé en mémoire.

Ces paradigmes indiquent que l'attention et la fovéa sont bien coordonnées, mais qu'il ne s'agit pas de processus limitatifs, à savoir que la prise d'information ne se limite pas à la zone fovéale.

✓ *Le regard et l'attention visuelle partagent des composantes neuronales communes*

Les études en neurophysiologie apportent leurs pierres à l'édifice pour relier les processus oculomoteurs et l'attention visuelle. En effet, les réseaux de neurones gérant les activités des mouvements des yeux possèdent des propriétés leur permettant de supporter l'attention visuelle ([Bisley03], [Thompson05]). Le réseau oculomoteur est dispersé sur plusieurs zones du cerveau : champ visuel frontal, zone intra pariétale latérale, colliculus supérieur et base de la formation réticulaire ([Mazer03], [Li-Z02], Figure 10). Le champ frontal visuel et le colliculus sont cruciaux dans ce réseau car ils reçoivent les signaux de nombreuses zones corticales et les transmettent à la base de la formation réticulaire [Sparks02]. Ces deux structures contiennent trois types de neurones, chacun identifiés par les événements spécifiques qu'ils encodent et produisent. Les neurones visuels encodent l'apparence d'un stimulus visuel, les neurones moteurs produisent l'initiation d'un mouvement saccadique vers une zone particulière de l'espace, et les neurones visuomoteurs encodent l'apparence d'un stimulus visuel et produisent le mouvement saccadique de l'œil [Munoz00].

Les systèmes d'orientation du regard et de l'attention visuelle semblent partager des composantes communes. En effet, les neurones du système oculomoteur encodent les informations visuelles sans qu'il n'y ait d'informations propres au mouvement d'encodées. De plus, les lésions de ces structures oculomotrices produisent un déficit de sélection attentionnel [Mesulam90] alors que la stimulation électrique de ces régions facilite la détection d'objets par l'attention [Moore04]. Enfin, ces structures reçoivent les informations depuis la voie visuelle centrale qui fournit les signaux permettant d'évaluer la saillance relative d'un objet [Schall97], donc du guidage ascendant de l'attention (processus bottom-up).

Une observation neurophysiologique complémentaire montre que les signatures neurales des processus oculomoteurs d'inhibition au retour et des processus attentionnels d'évaluation de la saillance peuvent être le continuum d'un même processus neuronal [Fecteau06].

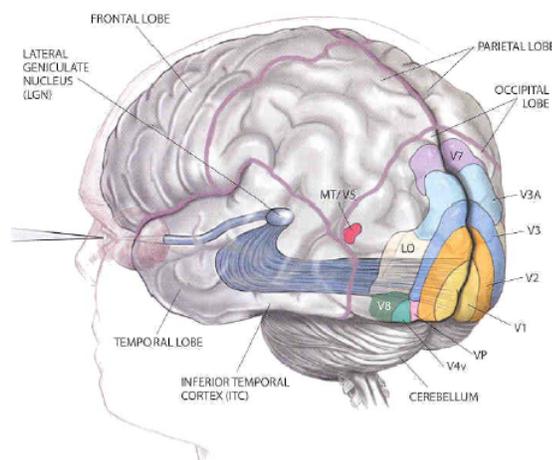


Figure 10 Schéma des zones neurales : le flux visuel part des yeux vers le noyau géniculé latéral puis vers le cortex visuel primaire, qui se trouve dans le lobe occipital postérieur [Rojas05]

Ces éléments permettent de comprendre qu'attention et comportement du regard sont intrinsèquement liés, mais que les processus qui les lient sont complexes et multiples. L'interprétation du comportement du regard est donc soumise à la plus grande prudence. Si attention et regard sont synchronisés, il est nécessaire pour aller plus loin dans notre souhait d'exploiter les mesures du comportement oculomoteur d'identifier les différents processus qui guident le déploiement de l'attention sur la scène, et qui peuvent se manifester par les processus oculomoteurs.

3.2.2 L'attention visuelle peut être guidée par des facteurs endogènes et exogènes

L'étude de l'attention visuelle est indissociable de l'étude du mouvement des yeux, et réciproquement. En analysant les facteurs qui influencent et contrôlent le regard, nous en déduisons les informations visuelles potentiellement traitées par le participant pour résoudre la consigne qui lui incombe. Les facteurs tels que les connaissances, les buts de la tâche, mais aussi les propriétés visuelles des objets, les formes et les contrastes, jouent des rôles importants dans l'orientation des processus attentionnels. Il y a ainsi des interactions complexes entre des processus de nature endogènes, c'est-à-dire provenant du participant (ses connaissances, son but de tâche ses intérêts, sa motivation, etc.) et des processus de nature exogène, c'est-à-dire provenant de facteurs extérieurs au participant, à savoir la représentation physique de la scène.

L'attention endogène résulte des spécificités propres au participant. Ici, le guidage attentionnel est alors appelé « top-down », c'est-à-dire que les lieux de fixations sont sélectionnés selon des indications fournies par le système cognitif. Les mouvements des yeux sont alors principalement contrôlés par les buts de tâche interagissant avec la nature sémantique de la scène et la mémoire ([Hayhoe05], [Henderson04a]).

L'attention exogène consiste à l'inverse en un guidage du regard par le signal visuel, c'est-à-dire que le flux informationnel montant (« bottom-up ») va orienter le regard. On parle alors de processus de saillance visuelle. Le contrôle du regard est ici une réaction automatique, rapide, et issue d'un traitement parallèle des propriétés visuelles du stimulus.

Ces deux approches ne sont pas antinomiques mais complémentaires. Nous verrons par la suite que selon le type de scènes, de consigne, et le déroulement de la séquence, ces deux types de guidage interagissent.

Dans notre travail, le terme de « *saillance* » est utilisé pour décrire les processus attentionnels « bottom-up » permettant de distinguer un objet dans la scène en fonction de ses caractéristiques physiques de celles des autres éléments de la scène. Le terme de « *relevance* » décrira quant à lui les processus attentionnels « top-down » permettant d'orienter le regard en fonction des buts de tâche ou des connaissances du participant. Le terme de « *pertinence* » renvoi aux processus attentionnels faisant émerger le sens sans relation à la consigne.

La transposition de cette approche dans le domaine acoustique permet de bien comprendre ces différents processus. La saillance physique est le processus orientant l'attention sur un bruit de marteau-piqueur émergeant dans un environnement calme ou d'un sifflet d'arbitre dans un stade. Dans le cas du marteau-piqueur, c'est la différence de puissance sonore en décibel entre l'environnement et l'outil qui crée l'orientation de l'attention ponctuellement sur la source. Dans le cas du sifflet, c'est la fréquence sonore du signal qui émerge de celle du bruit ambiant. La nature physique du signal a produit une orientation attentionnelle sur une source spécifique. La *relevance* cognitive est le processus orientant l'attention. Par exemple à partir des pleurs d'un nourrisson ou à partir d'un flux vocal d'un interlocuteur dans une foule. Le cas du nourrisson est associé à un conditionnement réflexe, faisant que l'attention détecte aisément un pleur de par son implication (le nourrir par exemple) et non parce que ce son était significativement différent du bruit ambiant en termes de signaux. De même, être capable d'échanger avec un interlocuteur dans un fort bruit ambiant produit par d'autres paroles signifie que l'attention permet d'analyser uniquement une source de sons, alors même que sa puissance sonore et la forme du spectre du signal n'est pas différent de celui de l'environnement (masquage attentionnel). Les deux processus sont parfois en interdépendance. Prenons l'exemple du moustique. Le bruit d'un moustique dans une chambre réveille parfois son occupant, alors même que sa puissance est très faible. Cependant, la fréquence et la constance de ce son associé aux conséquences douloureuses de la présence de l'insecte en font à la fois un son *saillant* physiquement et *relevant* cognitivement. L'attention est donc probablement sollicitée par les deux facteurs simultanément (saillance + *relevance*).

Le modèle ci-dessous (Figure 11), proposé par Navalpakkam ([Navalpakkam05] [Itti05]), est l'un des modèles synthétisant les différents types de processus de guidage de l'attention (*relevance* cognitive et saillance perceptive), et donc sur la planification des lieux qui seront sélectionnées par le regard. La scène visuelle est décomposée en différentes couches de signaux : luminance, chrominance, puis les orientations, contrastes, saturations locales sont extraites (low level feature detector) puis pondérées pour fournir la carte de saillance. En parallèle, une reconnaissance du résumé sémantique de scène (« *gist* ») est effectuée, et sa caractérisation spatiale fournit le « *layout* ». Enfin, les effets de la tâche et des connaissances préalables sont pris en compte pour pondérer la carte *layout* et la carte saillance. La confrontation de ces différents facteurs permet de calculer la zone de l'image qui aura la plus forte probabilité d'attirer le regard de par ses propriétés physiques et sémantiques.

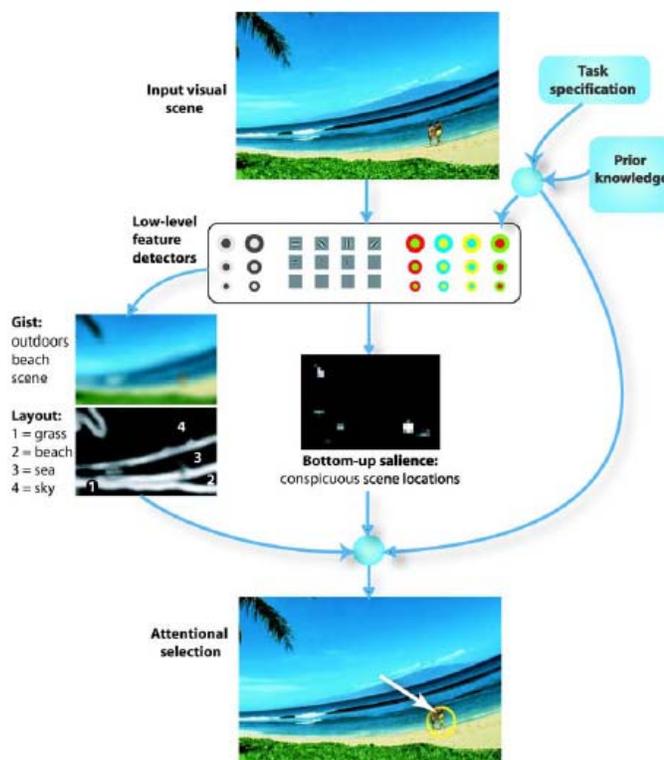


Figure 11 Modèle de déploiement de l'attention intégrant sur la base de facteurs de guidages physique et cognitif [Navalpakkam05]

Les informations visuelles de luminance et de chrominance sont séparées ; puis sont établies des cartes complémentaires extrayant les zones de contours, de discontinuités, de contrastes, etc. pour chaque carte sont dégagées les zones les plus attractives (zones de plus fort contours, de plus grandes discontinuités, etc.) et son recombinaison en une seule carte de prédiction. Par une voie parallèle, la signature basse fréquence de l'image est extraite, permettant de dégager les principales zones spatiales de l'image. Enfin, ces deux voies se rejoignent en une carte de saillance estimant la probabilité des zones locales d'attraction de l'attention exogène. Ce modèle intègre enfin une pondération des différentes zones de la carte de saillance par les facteurs de tâche et de pré connaissance, pour finalement extraire la zone qui aura la plus forte probabilité d'être traitée par le regard, donc par l'attention.

Ce modèle indique qu'il existe différents niveaux de traitement de l'information visuelle, et différents sens de transit de ces informations. On va ainsi s'intéresser aux différents processus influençant le déploiement de l'attention visuelle sur la scène. On présentera tout d'abord les processus d'origine exogène guidant l'attention visuelle, puis les processus d'origine endogène.

3.2.3 Orientation de l'attention visuelle par des facteurs de saillance visuelle

La complexité du monde visuel qui nous entoure produit un flux d'informations bombardant notre rétine à un débit de 10^8 bits par seconde, dépassant les capacités de traitement du cerveau humain [Tsotsos90]. Afin d'extraire les informations pertinentes, l'attention doit être dirigée sur certains éléments au détriment des autres [James90]. L'une des deux voies de guidage attentionnel est le facteur de saillance. La modélisation de ces processus consiste à prédire quel est le prochain objet qui

sera sélectionné par le système visuel, en se basant sur une carte topographique de l'espace établie sur des facteurs de distinction physique des objets.

Le guidage attentionnel par saillance physique a été tout d'abord observé au cours de tâches de détection de cible sur des représentations non réalistes (par exemple détection d'un élément rouge parmi un ensemble d'éléments verts) où les informations visuelles sont traitées de manière pré-attentive et en parallèle⁸. Les modèles computationnels du traitement du signal visuel ont été développés en utilisant les propriétés du système visuel, et plus précisément du cortex visuel primaire, afin de générer des cartes de saillance [Itti00]. Le principe repose sur le postulat que les régions porteuses d'informations ne sont pas uniformes en termes de signaux visuels car comportent des objets. Ainsi, les zones dont les variations perceptives sont suffisamment structurées seraient porteuses d'information. Par exemple, sur une scène de désert avec un chameau dans le sable, la présence de l'animal augmente la densité locale de contour, et de ce fait attire l'attention. Le déploiement de l'attention sur la scène, et donc la distribution spatiale du regard, serait alors corrélé avec ces pics informationnels [Oliva03].

Ces modèles de prédiction attentionnelle par analyse des caractéristiques des signaux constituant la scène visuelle reposent sur le concept de cartes de saillance, qui est une représentation bidimensionnelle de la scène permettant d'identifier et d'isoler les propriétés des signaux qui sont perceptivement pertinents.

⁸ Le traitement est parallèle lorsque le délai de réponse ne dépend pas de la taille de la scène ou du nombre d'éléments - [Wolfe94]

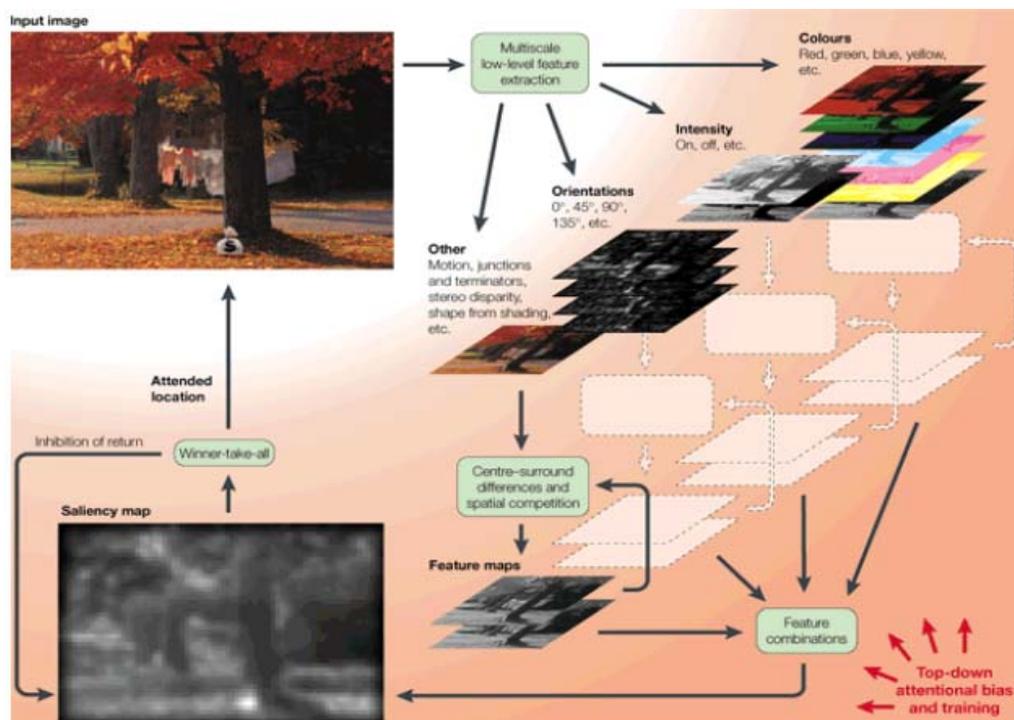


Figure 12 Principe de la carte de saillance, selon Itti & Koch : décomposition de la scène visuelle selon les dimensions de couleur, de contraste, de luminance, d'orientation des contours et discontinuités, puis calcul pour chacune de ces dimensions des zones de saillance, et synthèse de l'information sur une seule carte par combinaison [Itti01]

La carte de saillance est construite (Figure 12) en décomposant l'image selon ses dimensions visuelles physiques : cartes de chrominance, luminance, contours, intensité, orientation des formes ([Parkhurst02]; [Torralba03]; [Wolfe04]). Les cartes sont générées pour chacune des dimensions. Puis en fusionnant ces cartes par combinaison linéaire et en les pondérant selon une hiérarchie intrinsèque à chaque couche, le modèle génère des prédictions quantitatives sur les lieux de fixations et sur leur séquence (résultat spatial et ordinal). Le point le plus saillant sera considéré comme celui de la première fixation. Cette zone est ensuite inhibée, le calcul est relancé, le nouveau point émergeant est considéré comme celui de la seconde fixation, et ainsi de suite. Ce type de modèle est utilisé pour prédire les lieux vers lesquels l'attention exogène va être orientée, c'est-à-dire en l'absence de toute consigne ou de reconnaissance sémantique. Afin d'obtenir un modèle ordonné, une règle de « winner take all » (gagnant accaparant le tout) est appliquée. Cela signifie que seul le point de saillance le plus élevé est sélectionné, inhibant tous les autres points en compétition. Cela permet d'effectuer un ordonnancement des fixations. Une deuxième règle adoptée est l'inhibition au retour, qui diminue la saillance d'une zone qui a précédemment été gagnante. Cette règle se base sur des propriétés réelles du système visuel [Hooge00].

Dans un contexte écologique, la saillance physique est aussi associée au mouvement. Dans une situation dynamique de disparition/apparition soudaine d'un élément visuel, les distracteurs en mouvement dans une scène sont perçus rapidement et attirent l'attention. Il s'agit probablement d'un processus réflexe permettant la détection de dangers [Brockmole05b]. Les mouvements doivent ainsi être pris en compte pour le calcul de la saillance physique [Koch85].

Dans certaines situations, ces modèles prédisent correctement les fixations humaines [Parkhurst02], mais dans d'autres situations, l'efficacité du modèle est fortement réduite [Henderson06]. Comment expliquer ces différences ? Les réussites se différencient des échecs par le

type de scène et le type de tâche. Le niveau de représentation sémantique de la scène (abstraite *versus* concrète) et la tâche demandée au participant (active *versus* passive) semblent définir les limites de validité de la méthode. Ainsi, dès lors que l'essai s'effectue sur des scènes comportant des entités sémantiquement significatives et sous une tâche active (c'est-à-dire n'étant pas une exploration libre), les zones de fixations semblent très peu expliquées par les prédictions de saillance ([Turano03]). Certaines études suggèrent que les effets de la saillance visuelle ne peuvent être observés qu'uniquement lors de la phase d'identification de la scène (pendant l'élaboration de la significativité de la scène), et qu'ils diminuent avec le temps, au profit du guidage cognitif ([Henderson 99b]). D'autres auteurs montrent que même immédiatement après la première fixation, le regard peut déjà s'orienter en direction d'une cible recherchée, et non vers une zone de forte saillance physique, et cela que la cible soit présente ou non (c'est-à-dire que la scène ait été reconnue et la position vraisemblable de la cible identifiée- [DeGræf90]).

L'efficacité de la collaboration des processus de reconnaissance de scène et de mémoire semble donc responsable de la perte du guidage attentionnel exogène. En extrayant le résumé de scène (contenu lexical - « gist ») et le schéma de scène (configuration structurale - « layout ») lors de la première fixation, le système cognitif anticipe les lieux porteurs d'informations ([Brockmole06], [Castelhana05]).

Même sur des scènes abstraites, la saillance physique peut être modulée par le but du participant. Ainsi, lorsqu'il est demandé à des participants de détecter des singletons ayant une forme et une couleur particulière, les éléments qui ont la même forme mais qui n'ont pas la même couleur sont évités par les mouvements oculaires ; il y a donc inhibition perceptive des signaux non pertinents pour la tâche.

Ces éléments pourraient laisser penser que le guidage attentionnel par les processus de saillance se limite à des situations d'exploration libre sur des scènes abstraites. Cependant, il est observé sur des scènes réelles complexes et pour des tâches actives que les zones fixées contiennent, relativement aux autres zones de l'image, plus de hautes fréquences spatiales ainsi qu'une plus grande densité de contours [Mannan97b]. Complémentairement, le contraste local semble plus élevé et l'intensité plus faible sur les zones de fixations que sur le reste de l'image [Parkhurst03]. Il existe donc bien un lien entre saillance physique et attention visuelle, même s'il n'est pas possible dans ce cas de dissocier les effets de la saillance perceptive des effets de la présence d'informations produisant par leurs représentations visuelle des entités saillantes.

La principale difficulté d'application des modèles de saillance sur des scènes réelles provient de la multiplicité des facteurs qui influent sur l'attention. L'attention peut en effet se porter sur des informations nouvelles ou inattendues, sur des informations qui nous intéressent personnellement, sur des informations pertinentes pour la tâche, ou encore sur des zones qui comportent vraisemblablement une information attendue. Ces facteurs endogènes, sont liés à la connaissance, à l'expérience, à l'état d'esprit, aux préférences du participant, et déforment la carte de saillance au cours de l'exploration visuelle. Afin de pouvoir pondérer cette approche de saillance purement exogène par une approche liée aux connaissances, nous allons aborder la problématique de l'effet des processus endogènes sur le déploiement de l'attention sur la scène visuelle.

3.2.4 Orientation de l'attention visuelle par des mécanismes cognitifs

Les connaissances acquises au cours de la vie permettent d'identifier très rapidement une scène et son contenu, et peuvent orienter le regard dès les premiers instants de l'exposition ([DeGræf90], [Henderson 99b], [Loftus78]). Ainsi, Castelhana ([Castelhana03]) montre que les participants qui subissent une brève pré-exposition (permettant d'identifier la scène) trouvent plus rapidement une cible prédéfinie par rapport aux participants n'ayant pas subi cette pré-exposition. Leur regard se

positionne dès la deuxième apparition de l'image sur les zones pouvant contenir la cible. L'identification de la scène se ferait dans les toutes premières fixations, et les mouvements oculaires suivants auraient pour but de collecter des détails sur l'image [Rayner98].

3.2.4.1 Les buts de tâche ont un effet sur les mouvements du regard

Dans une étude référence de 1967, Yarbus [Yarbus67] montre que la perception des scènes réelles produit des trajets du regard (« pattern ») constitués de fixations et de saccades aux caractéristiques particulières. Le dispositif expérimental consiste à poser différents types de questions aux participants (Figure 13) alors qu'ils observent une peinture de Repin. L'auteur observe que les zones fixées par le regard sont influencées par la question posée. Les distributions des zones de fixations ne semblent pas s'organiser aléatoirement, mais plutôt vers les objets de la scène potentiellement informatifs selon la consigne. L'influence des demandes de la tâche sur l'exploration visuelle d'une scène est ainsi mise en évidence, suggérant que le caractère informatif d'une zone dépend de la consigne.

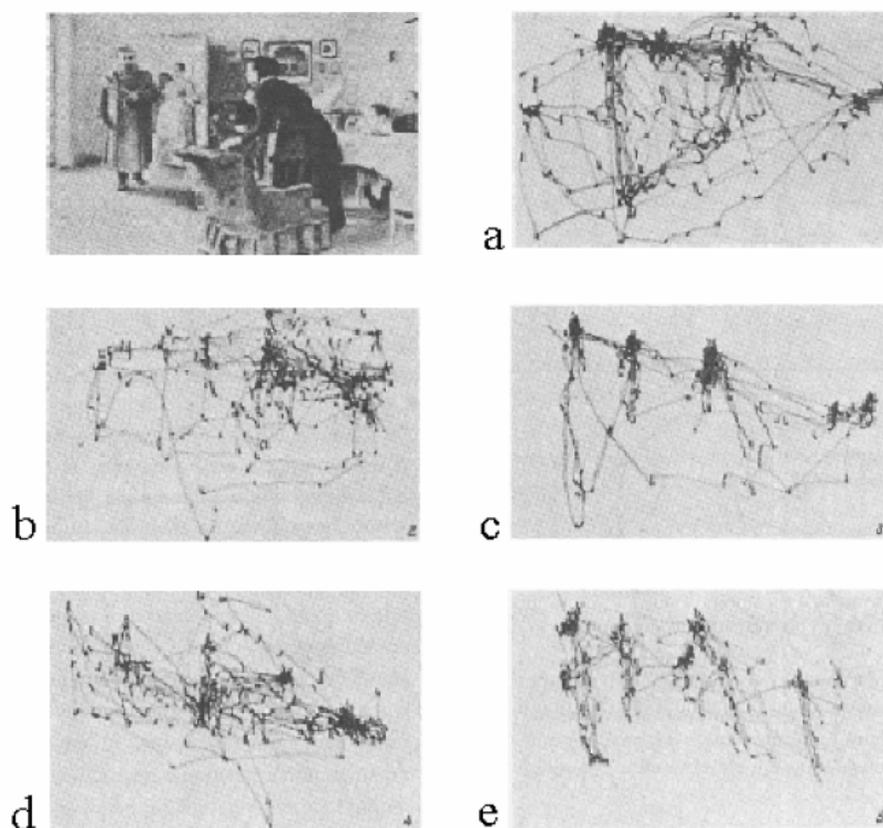


Figure 13 Peinture de Repin (1884-1888) « un visiteur inattendu ». a : exploration libre. b : estimer le niveau économique des personnages. c : juger leur âge. d : supposer ce que faisaient les personnages avant l'arrivée du visiteur. e : mémoriser les habits portés par les personnages [Yarbus67].

Mackworth et collaborateurs [Mackworth67] rapportent que bien que les mouvements des yeux d'un individu puissent être considérés comme lui étant propre, c'est-à-dire idiosyncrasiques, les zones sur lesquelles la distribution des fixations se construit semblent se corrélérer aux zones portant des informations «intéressantes». En exposant des photographies, ces chercheurs montrent que lors des deux premières secondes de l'exposition que dans les deux dernières, les fixations s'organisent autour

des zones informatives (ex: visage, texte, etc.). Buswell affirmait dans la même idée que « les mouvements des yeux sont des ajustements inconscients aux demandes de l'attention pendant une tâche visuelle » [Buswell35].

Cependant, parce que les yeux tendent à se poser sur des objets et non des zones vides et uniformes, il est difficile de distinguer les cas où (1) les fixations sont associées à l'importance de l'information pour la tâche, (2) les fixations sont associées à la présence d'objets quel qu'ils soient, (3) les fixations sont associées à des zones de forte saillance due à la présence d'objets. En effet, la présence d'objets produit une augmentation locale de la saillance physique, car présence d'information et saillance élevée sont corrélés. Les approches proposant que l'attention soit uniquement dirigée par le signal soit uniquement dirigé par la cognition se révèlent limitées. Il est donc nécessaire de prendre en compte de manière intégrative ces deux types de processus pour comprendre le déploiement de l'attention visuelle sur la scène.

3.2.4.2 La pré-connaissance du contexte influe sur l'attention visuelle

Si la consigne a un effet sur l'orientation de l'attention visuelle, et donc du regard, la mémoire exerce aussi une influence sur le traitement attentionnel des informations visuelles contenues dans la scène. L'un de ces phénomènes est dénommé « priming effect ». Notamment exploré par Friedman [Friedman79], cet effet est du à la reconnaissance contextuelle. Lorsque les propriétés structurales d'une scène correspondent globalement aux caractéristiques encodées en mémoire à long terme, l'attention visuelle va associer les informations locales aux informations mnésiques, influençant la représentation de la scène. Le phénomène d'extension des frontières (« boundaries extension ») en est l'une des manifestations [Bertamini05]. Lorsque l'attention est fortement guidée par la pré-connaissance de la scène, il va y avoir confusion entre les données mnésiques et les données perceptives. Par exemple, on expose aux participants dans une première phase une photographie de cuisine. Puis, dans une seconde phase, ces mêmes participants doivent reconnaître la photographie de la première phase parmi différentes photographies prises du même point de vue avec diverses focales. La plupart des participants choisissent alors une focale plus large que celle qu'ils ont vue au préalable. De par la pré-connaissance des objets présents dans une cuisine, ils ont enrichi leur représentation de la scène avec des éléments mnésiques. Un second phénomène d'orientation de l'attention par les connaissances est celui du « Contextual cueing » (ciblage contextuel) : sur les scènes réalistes [Brockmole06b] l'apprentissage par répétition de la position d'un objet cible en fonction du contexte oriente le regard dans la direction présumée de la cible dès que le contexte est reconnu. La modification des informations locales des scènes ne dégradent pas ce processus tandis qu'il disparaît si les informations globales du contexte sont modifiées. Brockmole en déduit que l'attention est orientée par le contexte, ce qui altère la détection d'un objet cible. Enfin, les paradigmes de cécité au changement permettent d'observer les situations dans lesquelles l'attention détecte ou non la disparition subite, temporaire ou non, d'un objet. Selon l'importance de l'objet pour l'identification sémantique de la scène, la détection de sa modification ou disparition sera plus ou moins rapide. Il y a donc bien un lien entre l'orientation de l'attention et la pré-connaissance de la scène.

3.2.4.3 La consigne influence les mouvements du regard

Tatler et collaborateurs [Tatler06] enregistrent 90000 saccades de participants réalisant deux consignes : exploration libre et détection de cible de luminance (zone de surbrillance gaussienne spatiale), sur 120 images de scènes naturelles. Dans la tâche d'exploration libre, 66% des saccades sont courtes (inf. à 8°) et dans la tâche de recherche 50%. Pour les longues saccades, quelle que soit la consigne, il n'y a que peu de différences de contenus fréquentiels des zones fixées par rapport aux zones non fixées. Cependant, pour les courtes saccades pour la consigne d'exploration libre, les zones fixées comportent significativement plus de hautes fréquences que les zones non fixées, donc des objets sont préférentiellement fixés. Les courtes saccades semblent donc associées à un processus

ciblant les hautes fréquences donc les objets. Cela pourrait signifier que pour la consigne d'exploration libre, les participants tendent à fixer les objets, à la différence de la consigne de recherche d'une cible de luminance, où les participants ne se basent pas sur les objets mais sur des propriétés « signal » de la scène. La possibilité que différents buts impliquent différents mécanismes de saccades a été suggérée par Frost et Poppel [Frost76]. Ces auteurs proposent que les saccades de plus de 10° sont exécutés par des mécanismes impliquant la voie du colliculus supérieur tandis que les plus petites saccades seraient générées par une voie géniculo-corticale. Si différentes voies sont impliquées, il est probable que différentes informations traitées [Bahill75]. Notons que dans des conditions de la vie quotidienne, les saccades courtes dominent [Land01b]. Cependant, Itti [Itti06] enregistre les mouvements oculaires de participants observant des films et classe ensuite les saccades en deux groupes (courtes et longues). L'analyse des propriétés statistiques de saillance sur les zones de fixation à destination des saccades ne montre pas de différences significatives entre ces deux groupes de saccades. On tire de ces expériences que les mouvements des yeux peuvent être dépendants du type d'informations recherchées.

Une autre approche de l'étude des propriétés des saccades et des motivations attentionnelles est effectuée par Velichkovsky, Unema et Pannash ([Velichkovsky05], [unema05], [Pannasch07]). Ces auteurs proposent qu'il existe deux types distincts de saccades : les saccades focales de faible amplitude et les saccades ambiantes de large amplitude. Ces saccades seraient associées à des durées de fixations spécifiques. Une saccade ambiante est d'amplitude supérieure à 5° et est suivie d'une fixation de durée inférieure à 200ms. Les saccades de plus faible amplitudes et suivies de plus longues fixations sont focales. Après avoir établis cette topologie, les auteurs proposent que lorsqu'une information locale est recherchée alors des saccades focales sont utilisées, et lorsqu'il n'y a pas d'orientation spécifique de l'attention, alors les saccades sont ambiantes. Les formes globales seraient ainsi traitées par de courtes fixations, afin de maximiser l'extraction d'informations diversifiées pour l'appréciation de la composition et du caractère global de l'image. Les détails focaux sont traités par de longues fixations, qui reflèteraient une exploration spécifique, dans le but de faire correspondre les objets représentés à un concept visuel sensé. Cette approche vient ainsi renforcer l'idée que le comportement du regard est guidé par des facteurs de recherche d'informations⁹.

3.2.4.4 Attention et variabilité inter/intra individuelle des mouvements oculaires

Noton et Stark répliquent et étendent l'étude de Yarbus [Yarbus67] sur l'analyse des mouvements oculaires sur des peintures représentant des scènes réelles ; ils montrent que les sujets tendent à fixer les « régions d'intérêt » selon un certain « scanpath », c'est à dire une stratégie d'exploration aux propriétés spatiales et séquentielles spécifiques ([Noton71a], [Noton71b]). Or, ces « scanpath » diffèrent selon les participants, ainsi qu'entre différentes répétitions (même tâche même scène) d'un même participant. Il existe donc certaines caractéristiques des mouvements oculaires qui sont communes à plusieurs essais. La collecte par Wooding et collaborateurs [Wooding02] de plus de 5000 séquences de visualisation de peintures sans consignes particulières apportent plutôt une réponse négative à l'homogénéité des trajets oculomoteurs. Ils installent en 2002 un oculomètre autonome au

⁹ Rucci [Rucci07] apporte cependant une proposition différente concernant les différences motivationnelles entre petites et amples saccades, courtes et longues fixations. Cet auteur propose que les petits mouvements oculaires soient associés à la présence de signaux de hautes fréquences. Pour une simple raison psychophysique, des mouvements de l'œil sont nécessaires pour percevoir finement les détails. Les prédictions des mouvements oculaires divergent alors de celles de Velichkovsky où des fixations focales sont plutôt de longue durée et de faible amplitude alors que pour Rucci, les fixations pour l'extraction de détails sont aussi de faible amplitude mais plutôt de courte durée.

National Gallery de Londres. L'analyse des données montre que seul un faible nombre de régions des peintures sont fixées communément par tous les sujets. Ce type de résultats avait déjà été présenté il y a de nombreuses années par Brandt [Brandt45] et Buswell [Buswell35]. Le premier auteur observa les mouvements oculaires respectivement dans la publicité et la peinture avec une consigne d'exploration libre. Il trouva de fortes différences interindividuelles dans les trajets oculomoteurs. Le second auteur a étudié les mouvements oculaires sur des peintures et photographies de plus de 200 participants et 55 images avec des consignes d'exploration libre et des consignes directives. Il rapporte que le comportement oculomoteur peut être classé en deux catégories, non pas en fonction des zones traitées mais plutôt en fonction du comportement du regard. Soit les participants effectuaient des successions de brèves fixations sur l'ensemble des zones principales de l'image, soit les participants effectuaient des longues fixations sur des petites sous régions de l'image. Cependant, dans les deux cas de figure, l'auteur observe que les zones d'intérêts traitées par les sujets comportaient principalement des éléments du premier plan de la composition, même s'ils étaient fixés dans des ordres différents. D'autres chercheurs tel que Groner [Groner84] ou Walker [Walker77] montrent sur des images de visages pour des tâches de reconnaissance et d'identification d'émotions qu'il n'y a que très peu de similarités dans la séquence temporelle des zones de fixations d'un essai à l'autre sur une image donnée pour une même tâche, pour un même participant.

Ainsi, si l'attention visuelle est effectivement corrélée à la prise d'information visuelle, elle est aussi guidée par de multiples facteurs exogènes et endogènes, tant liés à des spécificités de chaque participant qu'à la consigne ou aux propriétés signales de la scène. La prédiction des zones vers lesquelles se déploie l'attention ne peut donc pas se baser uniquement sur les informations pertinentes pour la consigne en prenant en compte un ensemble d'attracteurs attentionnels de différentes natures (synchroiquement et diachroiquement). Une approche statistique permet de limiter les effets de la variabilité idiosyncrasique de chaque essai mais risque d'être limitée pour modéliser le comportement du regard sur la base d'un unique facteur. Afin de mieux comprendre comment la perception et cognition interagissent dans une tâche d'évaluation visuelle, il est pertinent d'aborder la problématique des modèles intégrant à la fois des processus de guidage exogène et des processus de guidage endogène.

3.2.5 Modèles attentionnels intégratifs

Nous avons pu voir précédemment que les mouvements des yeux peuvent être guidés par des caractéristiques perceptives de la scène, c'est-à-dire que certaines propriétés des signaux visuels attirent l'attention, puis nous avons abordé les différents facteurs d'origine endogène qui peuvent orienter l'attention. Ces deux approches sont valides mais nécessitent d'être fusionnées pour expliquer au mieux le comportement du regard. C'est pourquoi nous abordons dans cette section quelques modèles intégrant ces deux approches.

3.2.5.1 Orientation du regard par détection d'informations sémantiques potentielles

Loftus et Mackworth [Loftus78] ainsi que Henderson et Hollingworth [Henderson97] ont montrés *via* des paradigmes impliquant des scènes naturelles que le regard se positionne plus facilement sur les zones comportant des objets pertinents tant perceptifs (contours, etc.) que cognitifs (objet incongru). De même, des études sur les enfants apprenant de nouveaux objets montrent que leurs regards suivent naturellement les zones de contours plutôt que de se poser sur des surfaces uniformes [Zinchenko62]. Il existe donc une tendance des fixations à se poser sur des zones potentiellement informatives ([Buswell35], [Mackworth67], [Henderson03]). Mannan [Mannan96] et Parkurst [Parkhurst03] montrent que les fixations tendent à se positionner sur les zones à haut niveau de contraste et de forte densité de contours (relativement aux zones ne recevant pas de fixations). Henderson [Henderson06] met en évidence dans une tâche de recherche d'objets sur scènes réelles

complexes de type « photographie d'intérieur de maison » que les algorithmes de saillance prédisent mal les fixations et leurs séquences, mais que l'intensité, les contrastes et les contours ont des valeurs spécifiques sur les zones fixées par rapport aux zones non fixées. Cela signifierait que les zones qui sont fixées ne le sont pas pour des raisons physiques, mais sémantiques, et que les zones porteuses d'informations possèdent des propriétés physiques spécifiques. Cet auteur observe de plus que les régions fixées sont notées comme plus informatives que les régions non fixées, et que les zones de textures uniformes ne reçoivent que peu de fixations. Ces zones considérées comme informatives sont donc localisées par des processus de vision globale car semblent fixées après leur contenu potentiel a été évalué. Il est cohérent que les objets significatifs doivent se dégager de l'arrière plan de la scène en termes de propriétés perceptives pour pouvoir être traités, et il est donc difficile de dissocier les effets d'attraction d'une zone par ses propriétés de saillance physique ou par la densité d'informations sémantiques, ces deux facteurs étant associés à d'importants taux de contours afin que l'objet se détache de l'arrière plan. Les zones fixées peuvent donc être attractives pour des raisons de saillance, mais leur intérêt peut être pondéré par la pertinence de l'objet pour la consigne. Ce serait ainsi les demandes cognitives d'informations qui orienteraient le déploiement du regard dans la scène [Henderson99] en pondérant les effets de la saillance visuelle, en tout cas pour les paradigmes de scènes réelles complexes.

Notons cependant que l'étude des facteurs orientant l'attention sur la scène doit intégrer deux limites. Dans la situation des scènes réelles, des objets sont disposés sur plusieurs plans (premier plan – arrière plan). Les objets de premier plan occultent en partie les objets de second plan. Pour une seule position de regard sur la scène, plusieurs plans peuvent être traités, donc plusieurs objets. L'information visuelle extraite n'est donc pas limitée à un seul item. La seconde limite est liée au concept de super-ordinalité des objets. Par exemple, lorsque l'on regarde une montre, analyse-t-on le cadran, les chiffres, les aiguilles ou l'ensemble fournissant l'heure ? Il n'est pas possible d'inférer de la mesure oculométrique le type d'information traitée. Remarquons que ces problèmes ne se posent pas en situation de lecture (où l'oculomètre est largement utilisé) car les mots sont clairement séparés par des espaces et qu'il y a nécessité d'extraire en sens à partir de l'ordonnancement plus ou moins linéaire des codes linguistiques utilisés.

3.2.5.2 Construction d'une carte de *priorité* combinatoire

Turano et collaborateurs [Turano03] ainsi que Henderson [Henderson 99b] observent que les modèles de saillance ne sont pas des modèles efficaces dans les tâches de vision active (avec consigne) sur des scènes réelles (naturelles ou intérieur de maison). Cependant, appliqué à des scènes représentant des formes géométriques simples, Yantis [Yantis93] montre que les variations des propriétés physiques d'un élément de la scène font varier son attractivité visuelle. Appliqué à des tâches d'exploration libre, Kahneman [Kahneman73] montre que le comportement du regard est contrôlé par les caractéristiques des stimuli telles que nouveauté, complexité, incongruité lorsque les participants sont invités à regarder la scène, sans autres formes de consignes. Or, ces facteurs sont tout autant perceptifs que cognitifs.

Le contrôle du mouvement des yeux serait donc soumis à un ensemble de processus complexes car il ne s'appuie pas uniquement sur les informations perceptives mais aussi sur la mémoire à court terme pour les informations précédemment traitées dans la scène, sur la mémoire à long terme dans le domaine spatial et sémantique concernant des scènes similaires déjà traitées, ainsi que sur le but et les intentions du participant. Enfin le contrôle du mouvement des yeux est soumis à différents facteurs en fonction du temps.

Afin de synthétiser ces interactions dans des modèles de prédiction, le concept de carte de *priorité* a été avancé ([Rao02], [Oliva03]). Une première proposition est d'établir une carte de saillance physique et une carte de relevance cognitive. La première carte représente les attracteurs perceptifs,

et la deuxième représente les attracteurs sémantiques au regard de la consigne à effectuer. Un processus dynamique produirait un remplacement progressif de la carte de saillance par la carte de relevance dans le guidage attentionnel, en fonction du temps d'exposition, c'est-à-dire que plus la scène est identifiée et analysée, moins les attracteurs perceptifs sont influents. Cependant, même la toute première saccade sur la scène peut être dirigée vers la position présumée d'une cible, qu'elle y soit présente ou non, impliquant qu'il n'y a pas nécessairement de délai à l'activation du guidage attentionnel par des processus top-down. Il est donc nécessaire de moduler ce processus de transition selon le type de tâche, selon le niveau de pré-connaissance de la scène, etc. Une seconde proposition est de construire la carte de saillance de base en pondérant les objets cibles selon leur relevance (i.e. pertinence pour la tâche, voir [Rao02]), selon le sens accordé par le participant en fonction de son implication et de sa motivation. Une troisième proposition est de combiner cartes de saillances et cartes de relevance sous forme d'intersection, c'est-à-dire que seules les parties saillantes des zones pertinentes émergent. Oliva [Oliva03] filtre ainsi une carte de saillance perceptive en utilisant une carte de pertinence des régions contenant vraisemblablement des objets cibles.

3.2.5.3 Positionnement du regard pour une optimisation de l'acquisition des informations

Une hypothèse indépendante des processus de saillance et de relevance provient de l'observation que le regard se pose de manière récurrente sur le centre de gravité spatial d'un objet ou d'une forme ([Denisova06], [Melcher99]) ainsi que sur les hauts contrastes locaux ([Mannan97b], [Reinagel99]). Kaufman [Kaufman69] montre que les participants posent ainsi leur regard à l'intérieur des angles formés par les fortes courbures de lignes, sans que celles-ci ne soient directement fixées. La sélection des zones de fixations pourrait donc être associée à une optimisation spatiale (maximisation) de la prise d'information afin de fournir les informations adéquates au bon moment ([Hayhoe05], [Land01]). C'est pourquoi plusieurs études récentes ont cherché à comparer directement les prédictions des théories de maximisation de l'information et les données oculaires ([Najemnik05], [Renninger05a], [Renninger07], [Raj05]). Renninger [Renninger07] part du constat de l'existence du processus d'inhibition du retour (voir fin du paragraphe), de l'anticipation du regard sur l'action [Hayhoe03], et du fait que les fixations traitent spécifiquement les zones de hauts contrastes [Raj05].

L'auteur considère alors qu'il est envisageable que les mouvements oculaires se positionnent naturellement sur des zones maximisant le traitement de l'information visuelle. La vérification expérimentale de l'hypothèse montre que les lieux des fixations dans des tâches de mémorisation ou de comparaison de formes abstraites correspondent à des positions depuis lesquelles la prise d'information est optimale. La comparaison des prédictions des modèles de saillance avec des modèles de maximisation de l'information [Renninger05b] lors de consigne de mémorisation de formes abstraites semble montrer que le modèle de saillance est moins performant. L'efficacité de cette approche dépend cependant du paramétrage du modèle en termes de coût de déplacement oculaire et de gain apporté par l'information extraite.

Lee et Yu [Lee00] confortent la plausibilité biologique de cette modélisation. Ils considèrent que la quantité d'information contenue dans une zone dépend à la fois des connaissances du participant, de la consigne, et du type de signaux physiques. Du fait des limites des résolutions du système visuel, le regard est déplacé de lieu en lieu selon un principe de minimisation de la dépense énergétique, donc de maximisation séquentielle de la collecte d'informations.

Pour expliquer le fait que pour la même consigne appliquée à la même scène, les mouvements oculaires ne suivent pas les mêmes séquences, un argument est que pour une même stratégie de recherche d'information, l'estimation de la qualité d'un centre de gravité informationnel est instable.

Le processus de maximisation de l'information interagit aussi avec des processus mnésiques. Ainsi, l'inhibition du retour [Posner85] est un processus qui diminue fortement l'attractivité d'une zone déjà fixée pour encourager l'exploration de nouvelles zones. Ce processus se vérifie autant en terme spatial

que sémantique [Leek03]. Ainsi, l'attention est modulée par les connaissances acquise dans un passé immédiat et influence directement le mouvement des yeux. Cela implique que malgré une grande richesse informative, une zone puisse être ignorée dès lors qu'elle aura été traitée une première fois. Ce phénomène confirme ici l'idée que la prise d'information soit optimisée pour une maximisation du traitement de l'information visuelle tant en terme de positionnement spatial qu'en terme de recyclage des données mnésiques.

3.2.5.4 Chronologie des processus attentionnels

Nous avons pu voir dans les sections précédentes que plusieurs facteurs peuvent se combiner dans le guidage du regard sur la scène. Pour prendre en compte efficacement cet effet, il est pertinent d'observer s'il existe une succession chronologique de ces processus de guidage. Il a été montré plus haut que l'identification de la scène s'effectuait dans les toutes premières fixations. Mannan, Ruddock et Wooding [Mannan96] montrent sur des images de scène naturelles pour une consigne d'exploration libre que les lieux de fixations dans la première 1.5 seconde d'exposition sont similaires pour une image réaliste et sa version issue d'un filtrage passe-bas. La majorité des objets n'étant pas visibles dans le cas de l'image filtrée, ces résultats suggèrent que le processus de sélection des premiers lieux de fixations ne se base pas sur d'identité des objets, mais soit sur la saillance, soit sur des éléments de type enveloppe spatiale permettant d'identifier la scène. Buswell et Yarbus ([Buswell35]; [Yarbus67]) avaient déjà observé que plus le temps d'exposition augmente, moins l'homogénéité interindividuelle des zones de fixation est bonne, indiquant que l'évolution chronologique des processus de guidage de l'attention diverge. Tatler [Tatler05] étudie l'homogénéité spatiale des zones de fixations au cours du temps entre participants ainsi que l'évolution des valeurs de la saillance visuelle des zones fixées. Il pose au préalable quatre hypothèses.

- La première est celle de divergence de saillance. L'équilibre entre les processus de saillance et de pertinence dans le guidage attentionnel évoluerait au cours du temps. Les processus bottom-up seraient plus influents au début de l'exposition et diminueraient progressivement [Parkhurst02]. Cela impliquerait une diminution de l'homogénéité inter individuelle au cours du temps, et que la saillance des zones fixées au début d'exposition serait significativement plus élevée que celles des zones observées en fin de session.
- Une deuxième hypothèse est celle de « salience rank » (ordre de saillance). Les lieux de la scène sont ordonnés selon leur saillance visuelle, et le système visuel sélectionne séquentiellement ces différents lieux. Puisqu'il y a vraisemblablement sur la scène peu de zones de hautes saillance (forte probabilité d'être fixées) et beaucoup de zones de faible saillance (faible probabilité de fixation pour chacune), l'homogénéité inter-participant des zones de fixations doit donc diminuer avec le temps en parallèle de la diminution du taux de saillance des zones traitées.
- La troisième hypothèse de « sélection aléatoire avec pondération spatiale » propose que les zones de fixation soient sélectionnées aléatoirement mais en respectant des niveaux de saillance et de relevance élevés [Melcher01]. En fournissant un point de départ commun aux participants, l'homogénéité inter sujet va diminuer avec le temps, tandis que le taux de saillance des zones fixées ne devrait pas spécialement évoluer.
- La quatrième hypothèse est celle de « divergence stratégique ». Les caractéristiques de saillance des zones fixées ne devraient pas évoluer avec le temps. Les stratégies d'exploration visuelle des sujets évoluent au cours du temps mais que les cartes de saillance n'évoluent pas. Cela prédit une augmentation de la variabilité des lieux de fixations au cours du temps mais pas de tendance des valeurs de saillance des zones fixées.

Les résultats de cette étude sur des scènes naturelles avec une consigne de mémorisation montrent que l'homogénéité des lieux de fixations entre participants diminue après les toutes premières fixations (environ 3 fixations). Les taux de saillance des zones fixées au cours du temps restent significativement élevés. L'hypothèse de divergence de stratégie est donc privilégiée. Il existerait ainsi une représentation intermédiaire constante de l'image issue d'une combinaison des cartes de saillances et des cartes de relevance. C'est l'interprétation et l'inspection de cette représentation qui évolue avec le temps et qui fait varier les zones considérées comme cible pour le regard.

La modélisation des processus attentionnel est un champ en pleine évolution, car les corrélations entre les fonctions neurologiques, les processus perceptifs et les activités cognitives sont de mieux en mieux spécifiées. On constate que si le regard et l'attention sont effectivement liés, il n'est pas évident de déduire du comportement oculomoteur les facteurs qui orientent l'attention et les informations visuelles acquises dans le cas de l'observation de scènes réelles. Malgré cela, il semble intéressant d'observer la manière dont se déploie l'attention sur la scène, afin de mettre en évidence, par une approche statistique, si les zones fixées possèdent des propriétés particulières tant sémantiques que physiques. Après avoir observé le lien entre le mouvement des yeux et l'extraction de l'information visuelle, nous allons nous intéresser aux processus de traitement des informations visuelles dans un but d'évaluation. Le chapitre suivant va présenter les principaux processus qui permettent l'évaluation visuelle d'un objet, puis les biais perceptifs et cognitifs qui interfèrent avec les informations recueillies par le regard. Nous chercherons à cerner les liens qui peuvent exister entre la mesure des mouvements du regard et les processus d'évaluation visuelle d'objets.

3.3 De l'identification à l'évaluation

Notre travail consiste à étudier les processus perceptifs et cognitifs impliqués dans l'évaluation visuelle du produit automobile. Dans ce contexte, nous nous intéressons à mieux comprendre comment se déroule une évaluation, et quels sont les processus qui pondèrent les facteurs de décision pour produire le choix.

3.3.1 L'évaluation par l'attitude

L'exposition d'un objet à un observateur engendre un ensemble de réactions. L'une des premières réactions est de type affectif, et émerge en parallèle de l'analyse objective de l'objet. Cette réaction de première impression oriente l'ensemble des processus perceptifs et cognitifs impliqués dans l'évaluation du produit. Il s'agit de l'attitude, qui est une réaction modérée à un stimulus spécifique¹⁰. De manière plus précise Fishbein et Ajzen [Fishbein75] définissent l'attitude comme « une prédisposition acquise permettant de répondre de manière cohérente, soit favorablement soit défavorablement, à un objet donné » (« a learned predisposition to respond in a consistently favourable or unfavourable manner with respect to a given object »). L'attitude est donc la « première impression » affective à un objet, émergeant d'expériences préalables et activée par un ensemble de stimuli, de favorable à défavorable. Russel [Russell91] propose complémentirement que l'expérience

¹⁰ Les réactions affectives sont classiquement classées en trois catégories : L'attitude, qui est une réaction modérée à un stimulus spécifique, L'humeur, qui est un état de sensations interne, modérées et durables, L'émotion, qui est un état de sensation interne intense

affective face à un produit s'organise en deux dimensions, dans lequel se positionne l'attitude : le premier axe est la dimension « plaisant-non plaisant », et le deuxième axe est la dimension « intérêt-désintérêt ».

La construction de l'attitude peut s'effectuer soit par une voie directe lorsque le produit est lui-même manipulé, soit par une voie indirecte lorsque l'individu a été soumis à de la publicité ou au bouche à oreille concernant le produit. Trois composantes d'antécédence peuvent influencer l'attitude : l'opinion sur le produit (« il est solide ; il est mal conçu »), l'émotion (« je n'ai pas aimé ; il est super »), et le comportement (« je l'ai déjà acquis ; il ne me servira à rien »). Selon le contexte, l'influence de ces trois facteurs va se pondérer différemment. Cet état mental qu'est l'attitude est particulièrement intéressant à observer car il peut être en lien avec le comportement du consommateur face au produit, aidant la catégorisation et la prise de décision. Cependant, il n'est pas possible de quantifier l'attitude de manière directe. Il faut donc l'inférer à partir de réponses, verbales ou non. L'utilisation d'échelles permet de recueillir des informations sur l'agrément du produit, mais ces notes portent plutôt sur la satisfaction en non pas directement sur l'attitude. La satisfaction serait la somme des ressentis négatifs et positifs issus de l'évaluation de différents attributs, donc directement influencé par l'attitude, mais pas exclusivement car aussi influencé par la confirmation des attentes (Bailey et Pearson (1983), Oliver 1993). Ainsi, l'attitude est une prédisposition de base (attitude favorable ou défavorable) issue d'une première impression, qui sera ensuite confirmée ou infirmée. Cette confirmation pourra produire des biais de traitements (surestimation, déformation perceptive, discrimination exagérée).

Ajzen [Ajzen85] étudie les facteurs qui influencent le comportement en plus de l'attitude afin de mieux corréliser attitude et comportement. Il pose alors la théorie de l'action raisonnée afin de prendre en compte l'influence d'autrui à la fois par la force de l'opinion d'une personne influente mais aussi par le poids accordé à cet avis sur le comportement découlant de l'attitude.

Après avoir posé l'idée que l'attitude est une composante fondamentale de l'évaluation du produit par l'initiation affective d'un mode de fonctionnement perceptif et cognitif, nous nous intéressons aux critères sur lesquels peuvent porter l'évaluation. En effet, si la « première impression » est automatique et rapide, et qu'il est demandé au participant une réponse analytique, il est pertinent d'envisager qu'une seconde phase, attentionnelle et séquentielle, se mette en place pour l'analyse du produit.

3.3.2 L'évaluation par les attributs

L'élaboration d'une opinion sur un produit est conséquente au développement de structures cognitives liant les attributs physiques du produit à sa représentation mentale. L'accès à l'information décrivant les propriétés d'un produit n'est pas direct car transite par une série de filtres perceptifs et cognitifs, interagissant notamment avec la pré-connaissance et le contexte. Deux types d'informations peuvent être employés pour établir la représentation et l'opinion sur un objet : les informations internes (informations encodées en mémoire au préalable), et les informations externes, perceptives, utilisées pour vérifier un critère de décision spécifique sur le produit, pour vérifier la correspondance entre les informations perceptives et les caractéristiques de la catégorie du produit, ou pour créer (ou renforcer) la base d'informations sur le produit pour un usage ultérieur. La recherche d'informations internes est notamment activée lorsque les participants rencontrent un nouveau produit et qu'ils utilisent leurs expériences préalables pour appréhender plus aisément la nouveauté. Beatty et Smith [Beatty87] définissent la recherche d'informations externes comme l'effort perceptif et attentionnel alloué à l'acquisition de données issues de l'environnement. Cet effort peut être utilisé en tant qu'indicateur de la quantité d'information acquise, du type d'informations traité, et de l'ordre d'acquisition des informations. Les informations externes sont portées par le produit lui-même, c'est-

à-dire par ses attributs. Selon Lancaster [Lancaster66] les attributs renvoient aux aspects physiques objectifs d'un produit. Grunet (1989, cité par [Keinonen98]) étend cependant ce concept à n'importe quelle caractéristique du produit permettant de le comparer à un autre. Les attributs sont alors :

- intrinsèques (physiques) ou extrinsèques (c'est à dire ajoutés par la suite - ex : prix)
- abordés individuellement ou globalement.

Lefkoff-hagius et Mason [Lefkoff93] classent les attributs en trois catégories :

- La première catégorie est celle des attributs de caractéristiques, c'est-à-dire des propriétés physiques d'un produit
- La deuxième catégorie est celle des risques et gains que le produit implique
- La troisième catégorie est celle des attributs d'images sociales (identité, valorisation, prestige, appartenance, différenciation sociale, etc.), qui décrivent la modification du détenteur du produit dans son contexte social

L'évaluation est issue d'un compromis entre les différentes notes positives et négatives d'un ensemble d'attributs. La pondération (l'importance) d'un attribut est déterminée par son effet sur l'identité du produit et dépend de chaque personne.

Myers et Alpert (1977, cité par [Jaccard86]) proposent que la pondération d'un attribut soit définie dans un espace à trois dimensions :

- La première dimension est celle de saillance/prégnance [MacKenzie86], c'est-à-dire la capacité d'un attribut à attirer l'attention de par sa nature cognitive ou perceptive
- La deuxième dimension est la valence, c'est-à-dire la note qu'obtient l'attribut en termes de "favorabilité" (positive ou négative)
- La troisième dimension est l'influence relative, c'est-à-dire la puissance avec laquelle l'attribut va influencer la prise de décision

Jaccard [Jaccard86] suggère de manière complémentaire que l'influence d'un attribut peut être vérifiée selon que sa disparition dénature le produit ou non.

Dans un modèle analytique (modèle décisionnel raisonné séquentiel, conscient, coûteux en énergie, et en temps, opposé aux modèles heuristiques), l'évaluation d'un produit nécessite donc d'analyser les attributs, pour ensuite de fusionner des différentes notes obtenues par chaque attribut afin de prendre une décision finale. Plusieurs règles cognitives aident à intégrer les différents attributs :

- La première règle est la règle de non-compensation. Elle permet de simplifier la prise de décision car la faiblesse d'un attribut ne peut être compensée par la force d'un second attribut. L'attention ne se porte dans ce cas qu'uniquement envers certains attributs spécifiques.
- La seconde règle est à l'inverse celle de compensation, qui s'applique lorsque l'implication du participant est forte. Si un attribut reçoit une mauvaise évaluation, son importance est alors considérée comme faible, et donc limite ainsi la diminution de la note globale.

Complémentairement, des règles de décision peuvent se développer au cours de la tâche de sélection d'information, favorisant les attributs situationnels (Payne et coll. [Payne93] - je découvre un attribut que j'aime beaucoup, je laisse de côté ma démarche analytique et non compensatoire au profit d'une réponse affective).

Cependant, la capacité de la mémoire de travail des objets (7+/-2 items) limite le nombre d'attributs visuels qui peuvent être appréhendés en parallèle¹¹. De fait, si toutes les propriétés des attributs telles que sa valeur et sa pondération doivent être traitées pour que puisse se former l'évaluation globale, 3 ou 4 attributs combleraient la capacité de la mémoire de travail. Afin de palier à cela, une stratégie d'évaluation par phase se développe, où chaque attribut est évalué un à un, et seul le résultat de l'évaluation est stocké en mémoire, permettant de traiter plus de critères que ce qu'il serait possible de faire par une approche parallèle. La prise de décision finale ne se fait donc pas uniquement sur l'évaluation de l'ensemble des attributs, mais aussi sur l'ensemble de résultats de sous séquences d'évaluation.

L'évaluation ne repose pas uniquement sur les attributs analysés attentionnellement car les participants confondent parfois la nature réelle d'un produit et la représentation de celui-ci [Braun99]. Il en résulte une attribution erronée des causes externes ayant produit un changement de l'état interne (cognitif et affectif). Ainsi, il peut y avoir une différence entre les éléments ayant influencé l'évaluation et les éléments que l'on considère avoir pris en compte pour l'évaluation, car des biais non identifiés auront impactés le traitement des informations. Ces biais peuvent intervenir à au moins deux niveaux : la perception et l'estimation. La perception d'un objet peut en effet être corrompue, car la familiarité, les avis préalables, ou les prévisions affectent la manière d'appréhender le monde. Ainsi, des processus d'inférences entre les différentes caractéristiques du produit [Broniarczy94] vont faire considérer qu'un attribut est présent car un autre attribut l'est (exemple : s'il y a un levier de vitesse en chrome, alors les sièges vont être vus comme recouverts de cuir : inférence de la sportivité). De même, la qualité perçue peut être construite sur la base des attentes plus que sur les caractéristiques réelles d'un produit [Hoch86]. Les attentes envers un produit sont aussi des facteurs d'erreurs. Lorsqu'une alternative (par exemple le choix d'un produit spécifique) est favorisée pendant le processus de construction du choix, l'information perceptive est biaisée en faveur de cette alternative. Après avoir effectué leur choix, les individus affirment avoir réellement perçue l'information sous la forme qui favorise l'alternative présélectionnée. Ce phénomène a été notamment observé dans le choix du consommateur [Russo98], ou les comportements de hiérarchie sociale en entreprise [Russo00]. La seconde catégorie de biais concerne l'estimation. Il a ainsi été observé que les participants, dans un processus d'ancrage, peuvent être amenés à exagérer les différences entre un produit qui sera choisi et un produit qui sera rejeté, afin de conforter leur choix [Svenson92] (biais de confirmation) ; ce biais s'active en général par la surestimation involontaire de la valeur d'un attribut [Hawkins90b]. Enfin, l'attitude initiale influence aussi l'estimation, sous la forme d'une illusion d'uniformité de la « première impression », c'est-à-dire que l'attitude initiale apparaît *a posteriori* comme homogène à l'évaluation et la valeur des attributs en accord avec la décision [Levine97]. Les hésitations initiales peuvent être occultées pour faciliter la prise de décision. Brenner, Koehler et Tversky [Brenner96] conduisent une série d'expérimentations en mettant en contexte un jugement de tribunal, pour lequel les sujets sont soit informés des arguments de la défense, soit des arguments de l'accusation, soit des deux parties. Il leur est ensuite demandé de prédire l'issue du jugement. Les observations montrent que les sujets exposés aux arguments d'un seul parti tendent à prédire que ce parti sera gagnant. Leur jugement est ainsi basé sur les arguments présentés, alors que les arguments du parti opposé, tout aussi pertinents, ne sont ni générés ni envisagés. Kardes [Kardes02] explique ce phénomène par un effet de survalorisation des arguments présentés en amorce, établissant une référence. Les individus vont chercher des preuves en fonction d'une hypothèse

¹¹ Le nombre d'items traité peut cependant varier en fonction du degré d'expertise du participant. Plus l'expertise est forte, plus on agglomère de l'information signifiante (Chunks de Miller [Miller60]) et plus on peut traiter d'informations. Cependant le seuil de saturation est atteint si on dépasse la règle générale, cette fois en nombre d'agglomérats informationnels.

première qui sera confirmée ou infirmée [Sanbonmatsu98]. Les éléments observés qui sont soit incohérents avec l'hypothèse première, soit confortant des hypothèses différentes, sont alors négligées. De forts biais apparaissent alors dans la sélection des informations et dans la décision, qui amènent à une conclusion prématurée en faveur de l'hypothèse première. Ce phénomène d'ancrage est défini par Sadler [Sadler73] comme un processus de « polarisation de l'attitude », c'est-à-dire que les participants testent l'hypothèse que leur première impression est effectivement correcte. Le traitement d'informations sélectives va conduire à confirmer l'attitude initiale car généré et intégré sous son influence. Les informations suggérant une différence seront négligées, résultant en une polarisation toujours plus extrême dans la direction initiale. Posavac, Sanbonmatsu, Kardes et Fitzsimons [Posavac05] observent ainsi l'existence d'effets de surévaluation positive des marques ou produits exposés individuellement. Lorsqu'une marque est évaluée de manière monadique, elle est jugée plus positivement que lorsqu'elle est jugée par comparaison. Un produit cible sélectionné au hasard parmi un ensemble de produits similaires est noté plus favorablement que la moyenne des valeurs d'appréciation issues du jugement de l'ensemble des produits du jeu duquel le produit cible a été tiré.

3.3.3 Effet du contexte sur l'évaluation d'un objet

Si l'évaluation dépend des attributs réels de l'objet, de la perception de ceux-ci puis de leur estimation, le contexte apparaît aussi comme influant la décision.

Le premier biais dû au contexte est celui de la manière dont une question est posée et le type de justifications attendues. Ainsi, Fitzsimons et Shiv [Fitzsimons01] mettent en évidence que lorsqu'il est posé des questions hypothétiques à un client sur un produit, son comportement se trouve impacté par l'orientation de l'hypothèse, alors qu'il avait été prévenu que la question était une simple hypothèse. Il a aussi été observé que l'évaluation du produit choisi est meilleure lorsque les participants doivent justifier leur choix, et la discrimination entre les différentes alternatives est plus grande [Simonson89].

Le second biais contextuel est lié au besoin de référence, car le jugement d'un produit est dépendant des autres alternatives, la décision s'élaborant par une adaptation automatique du contraste (des différences) entre les attributs de chaque alternative [Petty93]. Selon Sanbonmatsu [Sanbonmatsu98], la stratégie de jugement la plus précise est d'envisager toutes les hypothèses potentielles sur ce qu'est le meilleur choix (démarche hypothético-déductive), puis évaluer les attributs de chaque produit au regard de chaque hypothèse et alors prendre une décision basée sur les mérites relatifs de chacun. Cependant, du fait du coût cognitif que cela engendre et du délai temporel des processus, de nombreuses évaluations ne sont pas basées sur l'évaluation analytique ou sur la comparaison entre produits d'une même catégorie mais plutôt sur l'attractivité d'un attribut particulier d'un produit telle que l'image de marque, la qualité perçue, l'apparence... Ainsi, bien que la précision du jugement puisse être maximisée par la comparaison d'un objet spécifique avec tous les autres produits de la même catégorie, les consommateurs sont souvent en situation de choix dans laquelle un objet spécifique est évalué isolément (évaluation monadique). Le participant évalue les propriétés d'un objet en tentant d'estimer sa valeur propre en comparant connaissances et informations perceptives. La stratégie de traitement monadique est plus rapide et moins coûteuse cognitivement qu'une stratégie de comparaison dans laquelle de multiples itérations sont effectuées entre l'objet cible et les autres objets de la même catégorie ; il s'agit donc de la stratégie d'évaluation la plus courante tandis que l'évaluation comparative constitue plutôt une exception ([Dhar00] [Sanbonmatsu98] [Robinson85]). Cependant, l'évaluation sera moins précise et plus variable dans le cas d'une évaluation monadique par rapport à une évaluation comparative.

Ces éléments nous montrent que l'évaluation du produit est non seulement lié à ses caractéristiques, mais aussi de la manière dont les informations à propos de l'objet vont être diffusées.

3.3.4 Effets de l'implication et de l'expertise sur l'évaluation d'un objet

Le facteur humain est considéré ici sous deux aspects : l'implication du participant et son expertise dans la tâche ou sur le produit. L'implication est l'une des formes de lien entre un individu et un objet. Elle décrit le niveau d'activité du participant dans une tâche évaluative ainsi que la connaissance de l'objet par ce dernier. L'implication envers un produit est notamment définie par l'importance accordée à ce produit et la considération de l'enjeu de possession et d'achat en termes de statut et de « pouvoir ». Cette implication va permettre une élaboration de classes de produit de plus en plus détaillée. L'implication est donc directement liée à la connaissance du produit. Park et collaborateurs [Park94] montrent que l'implication peut être quantifiée par des tests de connaissances objectives sur une classe d'objets.

Trois niveaux d'implications sont en général considérés, pour trois stratégies d'évaluation [Solomon94], [Peter93]. Une implication peu élevée dans l'évaluation produit une faible motivation de collecte d'information sur les différentes caractéristiques du produit et sur ses alternatives. L'attention se porte alors sur un petit nombre d'attributs perceptifs facilement traités et la prise de décision s'effectue sur la base de connaissances plutôt superficielles (Olhavska et Grandbois 1975). Il s'agit d'une stratégie de réponse par routine. Lorsque les critères de décision sont établis mais pas la préférence entre ces critères, le nombre de critères de sélection diminue progressivement. Il s'agit d'une stratégie de résolution limitée pour une implication moyenne. Lorsque l'implication est élevée, une forte composante émotionnelle accompagne l'élaboration de l'opinion sur le produit ; un grand nombre d'alternatives et de valeurs d'attributs sont alors évalués. Le risque perçu de décision est ici élevé. Cette stratégie se retrouve aussi lorsque les critères de décision ne sont pas établis avant l'évaluation, ou encore que les produits ne sont pas familiers au participant. Le caractère d'implication ([Kapferer85], [Zaichokwsky85]) permet d'effectuer une discrimination entre des experts et non experts, entre personnes s'intéressant ou ne s'intéressant pas à un produit, entre ceux passionnés par un produit et ceux pour lesquels ce n'est qu'un simple outil.

Il existe d'autres caractéristiques que l'implication pour distinguer experts et novices en situation d'évaluation de produits. Les novices ont non seulement besoin d'acquérir plus d'information que ceux qui ont une pré-connaissance du produit, mais leurs capacités à exploiter les nouvelles informations sont aussi plus limitées car ils manquent d'outils cognitifs appropriés, de règles spécifiques à la situation (heuristiques), ou de stratégies de décisions, permettant de catégoriser correctement les nouvelles informations [Urbany89]. Ainsi, le type d'attributs et le choix des sources d'informations dépendront de l'expertise. Le développement de l'expertise engendre le passage d'une stratégie d'évaluation propre (ressenti à propos de) vers une stratégie basée sur les attributs. Les attributs qui sont immédiatement perçus par les experts ne le sont pas nécessairement par les novices, tout au moins pas avant un certain laps de temps. Les experts sont ainsi reconnaissables au fait de baser leur évaluation sur un ensemble varié d'attributs tandis que les novices le font sur leur ressenti et les recommandations d'autrui [King94]. Afin de limiter l'énergie nécessaire à l'évaluation, des participants novices peuvent employer des stratégies de contournement. Ainsi, la stratégie d'imitation consiste à adopter les décisions d'autrui, qui peuvent être des experts officiels (exemple : médecin pour un traitement), des leaders d'opinion, ou encore un vendeur conseillant un produit. Une autre stratégie de simplification est de baser la décision sur des attributs extrinsèques tels que la marque ou le prix. Feldman et Lynch [Feldman88] proposent enfin que l'accessibilité (niveau d'aisance à retrouver une information en mémoire) et la "diagnosticité" (niveau nécessaire et suffisant d'information pour effectuer un jugement) puissent être combinées de manière inconsciente pour simplifier le jugement.

Ainsi, plus l'effort effectué pour la remémoration d'information est grand, plus le jugement sera affecté négativement [Menon01].

Dans le domaine de l'art, la préférence esthétique diffère fortement selon les individus. Il a été proposé [Hekkert96] que les experts tendent à contrer leur réponse initiale affective au profit d'une évaluation intentionnelle et réfléchie [Winston92]. Tandis que la réponse initiale automatique peut refléter la préférence envers la familiarité, le mode de réponse secondaire intentionnel est caractérisé par un effort favorable à la nouveauté. Les experts sont plus à même de rechercher de nouvelles informations que les non experts car ils possèdent une structure catégorielle plus fine et doivent fournir un effort cognitif plus faible pour classer de nouvelles informations ([Alba87], [Hekkert95]). Les différences individuelles d'appréciation dépendent ainsi de l'intérêt et de l'expertise [Tobacyk79].

De ce paragraphe nous tirons que l'évaluation d'un objet se base sur un grand nombre d'informations, dont le traitement sera propre à chacun, tant en terme de sélection d'informations qu'en terme d'exploitation. Ainsi, pour toute expérimentation doit être défini le type de participants, car les processus attentionnels et décisionnels seront dépendants de leur expertise, de leur degré d'implication, de leurs expériences préalables. Nous nous intéressons dans notre cas à l'évaluation visuelle d'objets manufacturés. Après avoir observé les différents processus sur lesquels se base l'évaluation d'un objet, nous nous attardons dans le paragraphe suivant à un phénomène qui peut se produire spécifiquement dans le cadre de nos paradigmes : le biais d'évaluation issu de l'aisance au traitement des informations.

3.3.5 Influence de l'aisance du traitement de l'information sur l'évaluation

3.3.5.1 Concepts d'aisance du traitement de l'information

L'aisance du traitement des informations est un processus qui renvoie à la facilité avec laquelle une information portée par un stimulus est extraite. Cette aisance se manifeste par la vitesse, l'effort et la précision avec lesquels l'information est traitée par le système perceptif et cognitif lors de l'évaluation du stimulus. Ce terme a été introduit par des psychologues à la fin des années 80 [Jacoby89] et décrit un processus très important dans notre problématique car il relie la perception à la décision, les deux processus s'influencent réciproquement. Deux catégories d'aisances de traitement de l'information sont identifiées :

- l'aisance perceptive, qui dépend de variables telles que le contraste visuel entre l'arrière plan et le premier plan, la durée de présentation du stimulus, ou sa complexité visuelle
- l'aisance cognitive, qui dépend de variables telles que la difficulté d'identification sémantique, les relations sémantiques entre le stimulus et les connaissances, la cohérence entre le stimulus et le contexte, ou encore la disponibilité de processus cognitifs pour traiter les informations.

L'auteur distingue aussi l'aisance objective de l'aisance subjective, qui apparaissent comme corrélés mais pas nécessairement linéaires:

- l'aisance objective se mesure par la rapidité de réponse et son exactitude
- l'aisance subjective correspond au ressenti de la difficulté de décision.

L'aisance du traitement de l'information perceptive influence le jugement cognitif et affectif, et l'aisance cognitive influence le jugement d'attributs perceptifs. Cependant, parce que nous ne sommes pas conscients de l'ensemble des facteurs qui construisent nos décisions et nos états affectifs, nous pouvons interpréter de manière erronée l'aisance perçue, en l'attribuant à la mauvaise source. La suite de ce paragraphe présente plus en détails la nature de ces interactions.

3.3.5.2 Confusion entre facteurs d'aisance au traitement des informations visuelles

Afin d'étudier les caractéristiques de l'aisance perceptive, [Fazendeiro02] demande à des participants d'observer des images variées telles que des dessins, des objets abstraits, ou des photographies, puis de fournir une appréciation. Cet auteur note alors que l'aisance perceptive influence directement l'évaluation de ces stimuli. Les variables contextuelles telles que la durée d'exposition ou la multi-exposition, et les variables intrinsèques telles que la clarté, le contraste, les symétries, ou la complexité du stimulus, orientent l'appréciation. En effet, le fait que différentes variables puissent influencer la facilité à percevoir le stimulus produit des confusions entre l'appréciation du stimulus et sa facilité de traitement. Les confusions ne se limitent pas à des interactions perception-jugement mais aussi à des interférences perceptives. Ainsi, différentes études ([Whittlesea90]; [Witherspoon85]) montrent que l'aisance perceptive résultant de précédentes expositions au stimulus peut être attribué par erreur à la durée ou clarté de présentation du stimulus. Concrètement, les participants qui ont vu une scène dans un essai antérieur vont, lors d'une deuxième exposition, surévaluer la durée de l'essai ou sur-apprécier le contraste de l'image.

Une expérience de Masson et Caldwell [Masson98] consiste à faire effectuer par un groupe de participants une tâche de devinette, sous la forme « compléter la phrase ». Le groupe témoin n'effectue pas cette phase. Dans un deuxième temps, il est demandé à l'ensemble des participants de noter la clarté et la durée d'exposition de mots qui s'affichent sur un écran. Il apparaît que les participants qui ont été exposés dans la phase 1 à un ensemble de mots devinés et qui rencontrent de nouveau ces mots dans la phase 2 considèrent que la durée d'exposition était plus longue et la clarté visuelle plus élevée que les participants n'ayant pas subi cet amorçage sémantique. Il y a donc eu ici confusion entre l'aisance cognitive et l'aisance perceptive. L'aisance cognitive produite par l'amorçage sémantique a inféré le jugement d'une meilleure aisance perceptive.

Le processus inverse a aussi été vérifié : l'aisance du traitement de l'information visuelle peut être mal interprétée ou attribuée à une aisance cognitive. C'est ainsi que l'aisance à traiter les caractéristiques visuelles d'un stimulus peut être attribué par erreur à un sentiment de familiarité, et produire un effet d'illusion mnésique [Jacoby89b]. Une expérience de Whittlesea et collaborateurs [Whittlesea90] démontre ce phénomène. Dans une première phase, les auteurs exposent une liste de mots à des participants. S'ensuit une seconde phase, où des mots sont présentés sur un écran. L'aisance perceptive du traitement visuel des mots est manipulée par la superposition au mot de masques visuels composés de formes pointillées de densité variable. Les participants ont pour consigne de confirmer ou d'infirmer la présence du mot dans la liste de la phase 1. Les résultats montrent que les mots exposés avec le plus de clarté (densité du masque faible) sont plus vraisemblables d'être considérés comme appartenant à la liste initiale. Les participants confondent ainsi une forte aisance perceptive due à la clarté visuelle à une pré-connaissance, tandis que la réelle pré-connaissance avait peu d'effets. Cette illusion mnésique disparaît lorsque les participants sont informés de la manipulation de la clarté visuelle. Une étude de Fazendeiro et collaborateurs [Fazendeiro02] montre que l'illusion mnésique par manipulation de l'aisance perceptive peut être étendue à la manipulation d'aisances cognitive par des mots et des images. Ainsi, dans une première phase, les participants observent une liste référence qui inclut images et mots. Ensuite, les participants observent une liste test, en indiquant quels items étaient apparus dans la liste référence. La manipulation des items entre la phase 1 et la phase 2 est la suivante. Certains items étaient présents dans la référence et d'autres étaient nouveaux. De plus, certains des nouveaux items étaient non liés sémantiquement (NLS) avec la liste référence tandis que d'autres nouveaux items étaient liés sémantiquement (LS) à des items précédemment présentés. Ensuite, tous les items LS étaient présentés dans une différente modalité visuelle que celle de la liste référence. Par exemple, si le mot «bouteille» apparaissait dans la liste référence, l'item LS était une image de « verre de vin », tandis que l'item NLS était une image de « chaise ». Les résultats montrent que les participants

prétendent avoir vus les items LS, mais pas les items NLS. L'aisance avec laquelle ils traitent les items LS du fait de l'amorce sémantique est attribuée par erreur à une pré-exposition. La contextualisation sémantique a produit une aisance cognitive qui en retour a produit un sentiment de familiarité. Dans cette expérience aussi, l'illusion de mémoire disparaît lorsqu'il est dit aux participants que leur sentiment de familiarité peut être influencé par un facteur d'arrière plan, produisant ainsi une indétermination de la nature de la valeur informationnelle de l'expérience d'aisance [Winkielman00b].

La facilitation du traitement perceptif ou cognitif semble augmenter le sentiment de familiarité ([Klinger94]; [Smith98]). Or, les stimuli familiers sont souvent préférés aux moins familiers. Ces processus décrivent un effet formalisé par Zajonc en 1968 [Zajonc68], sous le nom de « mere exposure effect » (effet de simple exposition), qui, sur les bases d'une potentielle prédisposition biologique de la peur de l'inconnu, va lier aisance de traitement et niveau d'appréciation de la cible. Cet effet apparaît lors de l'exposition répétée à un stimulus. Il produit une augmentation graduelle de l'appréciation [Bornstein89]. Étudiée de nombreuses fois, notamment par Kunstwilson et Zajonc en 1980 [Kunstwilson80], la préférence par multi-exposition a été identifiée tant dans des cas d'expositions subliminale que pro-liminale, tant pour des percepts visuels sans signification que pour des scènes significatives. Ainsi, des formes géométriques vues plusieurs fois sont préférées à celles vues une seule fois lors d'un choix par paire. Puisqu'il n'existe aucune raison objective de préférer un polygone à un autre il a été considéré que la préférence est due au « mere exposure effect ». L'interprétation du phénomène se base sur le principe que l'augmentation du nombre d'expositions ou de la durée d'exposition cumulée améliore l'aisance de traitement tant perceptivement que cognitivement, par une analyse du stimulus facilitée par la reconnaissance. En retour, l'augmentation de l'aisance à traiter le stimulus influence directement l'appréciation même du stimulus, et ce pour deux raisons. Tout d'abord, l'aisance de traitement est en soi une expérience plaisante. Ce ressenti peut être considéré par erreur comme provenant de l'appréciation du stimulus. Ensuite, plus l'aisance est élevée, plus la suggestion de la familiarité est forte. Or pour des raisons d'adaptation, les stimuli familiers sont préférés aux non familiers. Ce raisonnement implique que ce n'est pas la multi exposition en soi qui oriente la préférence, mais plutôt la facilitation du traitement de l'information. Ainsi, n'importe quelle variable qui augmente l'aisance de traitement d'un stimulus pourrait influencer l'appréciation.

L'aisance perceptive influe sur la préférence, les participants appréciant favorablement les stimuli les plus aisés à appréhender perceptivement, quel que soit le stimulus employé et sa manipulation ([Reber98]; [Winkielman01]; [Winkielman00]). Cependant, cette influence ne se manifeste pas uniquement via des jugements explicites mais peut aussi être observée sur des mesures physiologiques. Des participants sont ainsi exposés à des images dont les informations sont plus ou moins aisées à traiter perceptivement. Pendant la tâche on effectue une mesure d'électromyographie faciale [Winkielman01], consistant à mesurer l'activité électrique des muscles zygomatiques. Une réaction positive à un stimulus produit un imperceptible sourire tandis qu'une réaction négative produit des froncements. Lorsque les images sont faciles à traiter, l'activité électrique des zygomatiques révèle une réaction positive, et négative avec des images difficiles à traiter. Cela implique deux éléments. Le premier est que l'aisance perceptive a un effet sur l'attitude, c'est-à-dire sur un état affectif ; le deuxième est que cette aisance se manifeste physiologiquement. Le modèle d'aisance affective propose donc que l'aisance de traitement puisse provoquer des émotions positives [Winkielman03], se basant sur l'idée que l'aisance indique une bonne reconnaissance d'un stimulus, une organisation cognitive cohérente, et la disponibilité de processus pour traiter la situation courante ; de telles processus tendent à être associées avec des émotions positives ([Carver81]; [Garcia00]; [Ramachandran99]; [Schwarz90]; [Vallacher99]).

3.3.5.3 *Synthèse*

Les observations de ce paragraphe sont importantes pour notre travail expérimental car elles mettent en évidence de nombreux mécanismes tant perceptifs que cognitifs qui sous-tendent le jugement, sans que l'attention ne soit directement impliquée. Tel que l'illustrent les influences de la pré-exposition et de l'amorce cognitive, les informations aisément traitées dépendent des pré-expositions individuelles, des connaissances, de la culture, donc du contexte. De plus des influences temporaires telles que la disponibilité de ressources attentionnelles ou la présence de centres d'intérêts dans la scène peut influencer l'aisance de traitement, résultant dans des variations systématiques inter et intra individus. L'appréciation par l'aisance peut être générée par la manipulation d'un grand nombre de variables issues du contexte ou du stimulus, perceptives ou cognitives [Ramachandran99]. Ainsi le jugement apparaît vraisemblablement influencé par de nombreux facteurs non liées aux propriétés de l'objet, sans qu'il n'y ait d'allocation de ressources attentionnelles sur ces variables. Les mouvements des yeux pourraient refléter ces processus d'aisance de traitements cognitifs et perceptifs des informations visuelles, sans qu'il n'y ait de liens entre les zones observées et le jugement.

L'aisance du traitement de l'information semble influencer l'évaluation en fonction des propriétés de l'objet, des ressources disponibles pour analyser le stimulus, et des expériences préalables. Le paragraphe suivant va analyser plus en détail les situations d'évaluation d'objets ayant un caractère typique ou nouveau, situations où l'aisance de traitement a un effet non négligeable sur les ressources attentionnelles et donc sur le comportement visuel.

3.3.6 Préférence, appréciation, nouveauté, typicité : des interactions « complexes »

3.3.6.1 *Typicalité versus nouveauté*

Il est reporté dans la littérature [Veryzer98] que la typicalité est un facteur influençant fortement l'évaluation d'un produit en général et l'appréciation esthétique en particulier. Dans le même temps, la nouveauté a été observée comme un facteur augmentant l'attractivité d'un produit [Simonson00]. Pourtant, la préférence des produits typiques semble s'opposer au désir de nouveauté [Whitfield79]. Comment concilier ces deux facteurs pour prédire la préférence des objets manufacturés ?

Le prototype est la forme dans laquelle s'exprime la typicalité d'un produit. Selon la théorie de la préférence pour les prototypes [Whitfield79b], plus un objet est prototypique, plus il sera apprécié. Cela s'observe tant pour des maisons, des designs intérieurs, des performances musicales [Repp97], que des produits de consommation [Veryzer98]. Barsalou [Barsalou85] propose que l'attribut de « tendance centrale » soit une distance entre un élément et la moyenne de l'ensemble des autres membres de la même catégorie (catégorie d'un même air de famille). Cet attribut est donc un déterminant important de la typicalité, où les résultats d'évaluation de la préférence esthétique montrent une très forte corrélation entre typicalité et tendance centrale. La notation subjective de typicalité reflète directement la note de similarité moyenne, donc la valeur de l'attribut de tendance centrale. La théorie du stimulus moyen [Langlois90] postule que l'exposition et l'expérience d'une certaine classe de stimuli au cours de la vie produisent des représentations moyennes pondérées et utilisés en tant que modèle prototypique. Plus un stimulus est proche de ce modèle, plus il sera considéré comme attractif. Ce processus a été observé tant sur les visages [Langlois90] que sur les objets comme les voitures [Halberstadt03]. Complémentairement, une grande distinctivité par rapport au standard moyen (prototype) est souvent associée à des jugements d'appréciation faible [Barsalou85], [Rhodes96]).

La typicalité et la nouveauté ne sont pas nécessairement des dimensions orthogonales; par exemple une lampe peut être innovante dans l'utilisation de nouveaux matériaux tandis que sa forme

et sa fonction restent familières. Hekkert [Hekkert06] montre ainsi que les deux facteurs, typicalité et nouveauté, sont aussi importants l'un que l'autre pour expliquer les préférences esthétiques pour tout type de produit et de participants (étudiants, non étudiants). Même pour un produit pour lequel l'attractivité esthétique peut ne pas être considérée comme une caractéristique importante (ex : ponceuse), le caractère novateur n'est pas moins prédictif de la préférence que le caractère typique. Cependant, la pondération de chacun de ces deux facteurs semble dépendre de propriétés individuelles, que pourrait expliquer le facteur « niveau d'expertise ». L'adhésion à un produit (son appréciation) peut ainsi être établie à partir de ses attributs de typicalité et de nouveauté. Cependant, ces deux facteurs sont non compensatoires, c'est-à-dire qu'ils déterminent l'appréciation d'un produit mais si l'un est négatif, il supprime les effets positifs de l'autre facteur.

Il existe ainsi deux mécanismes distincts influençant la préférence et l'attractivité [Hekkert03] explicables par une approche évolutionnariste [Bornstein89]. La préférence du prototype et l'attractivité de la nouveauté seraient deux effets issus du répertoire comportemental de base de l'individu. Ainsi, il est nécessaire à l'enfant d'être attiré par la nouveauté afin de favoriser son apprentissage et l'acquisition des connaissances. A contrario, pour un adulte, ayant tendance de limiter les risques inhérents à l'inconnu, il est plus sûr de préférer les items familiers, donc typiques. La préférence basée sur le prototype serait issue d'un traitement attentionnel automatique favorisant une classification rapide, ne nécessitant pas de conscientisation, et du moindre effort. Plus un produit correspond au prototype d'une catégorie déjà existante, plus l'adhésion est forte, l'attitude positive, et l'appréciation élevée. C'est ainsi que les effets du « mere exposure » se manifestent par une augmentation de la préférence lorsque la fréquence d'exposition au produit augmente, sans que l'on ait conscience (du fait de la nature automatique du processus) d'avoir vu le stimulus répété auparavant [Murphy93].

L'attraction pour la nouveauté, à l'inverse du moindre effort, est plutôt coûteuse. Ce processus se manifeste lorsque l'on est exposé à une forme atypique ou innovante d'un produit. Il est alors nécessaire d'inhiber les processus liés à la familiarité et d'activer le traitement de stimuli nouveaux. Se met alors en place un processus attentionnel et non automatique, favorisant la recherche de stimuli inconnus, et contrebalançant l'attitude favorable de la typicalité. Dans un registre comportemental, les consommateurs peuvent préférer des produits nouveaux car ils recherchent parfois la variété [Hutchinson86]. Les clients tendent aussi à sélectionner les produits uniques ou émergents [Van Trijp96]. De plus les meilleurs produits tendent à être plus cher, plus rares et atypiques [Veryzer98]. Cellierino [Cellierino03] montre que la théorie de la préférence pour les prototypes est parfois mise en défaut, car des versions féminisées de prototypes de femmes sont considérées comme plus attractives que le prototype lui-même.

Biederman et Vessel [Biederman06] avancent l'idée que des récepteurs opioïdes¹² qui sont dispersés dans le système visuel reçoivent plus d'opioïdes lorsque l'activité neurale est forte, produisant une expérience visuelle plaisante. En utilisant des scènes naturelles, les auteurs affirment que les scènes inconnues et fortement interprétables sont plus plaisantes à regarder car elles génèrent plus d'activité neurale dans des niveaux élevés, tels que le cortex inféro-temporal, notamment là où la concentration de récepteurs opioïdes est la plus forte. Dans cette approche, le système visuel agit comme une structure indépendante des comportements appétitifs.

On a pu voir dans cette section qu'il existe deux effets antagonistes qui influent sur l'appréciation des produits : la préférence pour les prototypes et la recherche de nouveauté. Selon les situations,

¹² Les récepteurs opioïdes sont les composants neuronaux dédiés à la fixation et la transformation des molécules opiacées

l'un prendra le pas sur l'autre. Il s'agit cependant de processus propre à chacun car renvoyant aux expériences préalables. Puisque la construction de l'évaluation interagit avec l'attention, cela peut expliquer que la prise d'information visuelle en situation de jugement puisse différer d'un essai à l'autre, d'un individu à l'autre, car les processus liés au prototype et à la nouveauté ne sont pas nécessairement statiques. En effet, le contexte lui aussi influe sur les effets de ces deux tendances.

3.3.6.2 Le contexte pondère les effets de la typicalité et de la nouveauté

Les facteurs contextuels tels que le temps d'exposition, l'état affectif du participant, ou l'environnement dans lequel un élément est présenté vont influencer la prépondérance des facteurs typicalité ou innovation sur l'évaluation. Ainsi, si la durée dont le participant dispose pour effectuer son évaluation est courte, la réponse fournie se basera sur les réactions les plus immédiates et automatiques, favorisant les stimuli typiques faciles à classer. Le facteur de typicalité sera alors un fort prédicteur de la préférence [Freund96]. Freund, Kruglanski et Bar-Tal's observent que l'exposition à haute fréquence de peintures concrètes produit l'attribution de notes d'appréciation plus élevées que lorsque l'exposition s'effectue à basse fréquence.

Un deuxième facteur contextuel qui peut influencer le poids relatif de la typicalité ou de la nouveauté est la « fonction esthétique » du stimulus [Hekkert96]. Les œuvres d'art sont conçues dans l'intention spécifique de produire une expérience esthétique, où l'activité cognitive attentionnelle est majeure. Le caractère innovant y est beaucoup plus intégré, et donc influent, que pour les objets qui ne sont pas perçus avec une attitude esthétique. Ainsi, Hekkert montre que si l'on spécifie aux participants d'évaluer la préférence esthétique d'un produit de grande consommation, la répétition des expositions a moins d'effets sur la note que lorsque la consigne ne porte pas sur l'esthétique. La typicalité (provenant de la multi expositions) a donc moins d'influence sur l'évaluation esthétique que sur d'autres types de jugements.

Les caractéristiques individuelles des participants peuvent aussi affecter la prépondérance d'un mécanisme sur l'autre (typicalité-automatisme *versus* nouveauté-attention). Il n'a cependant pas été démontré que le niveau d'expertise établi à partir d'un score de connaissances préalable permettait de discriminer d'une part une stratégie experte favorisant la nouveauté et d'autre part une stratégie novice favorisant la typicalité. Cependant, une autre forme d'évaluation de l'expertise qu'un test de connaissances pourrait peut-être permettre de mieux discriminer experts et non experts sur les effets respectifs de la nouveauté et de la typicalité dans le jugement.

Nous voyons dans ce paragraphe que l'appréciation d'un produit dépend de ses caractéristiques innovantes ou typiques, mais que le contexte d'évaluation va orienter l'appréciation vers l'un ou l'autre de ces deux attributs. La partie suivante décrit un paradigme expérimental permettant d'étudier les relations entre l'attractivité d'objets et leur caractère innovant en contrôlant la réaction répulsive que peut produire des stimuli trop innovant.

3.3.6.3 Innovation et attractivité

Le paradigme du « mere exposure effect » développé par Zajonc [Zajonc68] consiste à augmenter l'attractivité d'un stimulus par sa multi exposition, l'appréciation augmentant au fur et à mesure des répétitions. Ce paradigme manipule la fréquence d'exposition sans qu'il n'y ait de familiarisation attentionnelle avec le stimulus. Pourtant, l'implication du participant peut impacter l'évolution du jugement, il serait donc pertinent de familiariser le participant avec le stimulus avant qu'il ne l'évalue, augmentant par là même son implication dans la tâche [Crandall73]. La méthode RET: « Repeated Evaluation Technic » consiste ainsi à faire répéter l'exposition, et à pousser l'implication du participant dans l'évaluation. Carbon et Leder [Carbon05] proposent une mesure répétée de l'attractivité sur la base de ce « RET » afin d'observer l'évolution des effets de l'attractivité par la familiarité et l'attractivité par le caractère innovant sur le jugement.

L'une des applications du paradigme « R.E.T. » de Carbon et Leder consiste à faire évaluer le caractère innovant et l'attractivité d'un ensemble de dessins d'habitacles. Les représentations sont définies dans un espace à deux dimension : « innovation » et « courbure des lignes ». L'évaluation du caractère innovant et de l'attractivité se fait en deux étapes chronologiques, séparées par une phase d'analyse détaillée de chaque modalité visuelle via la notation d'un ensemble de caractéristiques.

Dans la première étape, les designs innovants sont considérés comme moins attractifs et plus novateurs. La typicalité et la faible distinctivité sont donc favorisés. Après la phase d'analyse détaillée, lors de la deuxième étape, les designs innovants deviennent non seulement plus familiers donc plus aisés cognitivement à appréhender. Les participants sont alors capables d'analyser la structure innovante du produit sans s'arrêter à une première réaction automatique. Ainsi, le design qui demande un traitement cognitif sophistiqué reçoit plus d'intérêt que le design peu innovant qui est aussi fait plus familier et prototypique (et donc plus ennuyeux). Ainsi, ce qui était innovant est rejeté dans un premier temps, mais ensuite préféré au prototype. L'atypicité et la forte distinctivité bénéficient donc de la répétition de l'évaluation mais aussi de l'analyse détaillée de chaque représentation (donc de l'implication). Cette inversion de tendance peut être due au fait que l'innovation limite l'ennui par sa nature sollicitante cognitivement, mais aussi que l'augmentation de la familiarité réduit l'incertitude [Lee01] et la menace [Stang74], permettant une analyse sans connotation négative. Remarquons que la note du caractère « innovant » de chaque modalité est la même entre les deux phases alors que la note d'attractivité évolue.

Une explication de ce caractère dynamique de l'attractivité se base sur le modèle de traitement par niveaux [Craik72]. Craik propose en effet que les informations portées par le stimulus soient traitées simultanément à différents niveaux cognitifs. Plus le traitement atteint des niveaux complexes, mieux il sera encodé en mémoire. Les caractères de nouveauté, d'atypicité et de distinctivité des designs innovants vont être élaborés à des niveaux profonds (car sollicitant des traitements basés sur le sens). Ils inscrivent donc plus durablement des traces mnésiques et sont moins susceptibles d'être oubliés. Les participants perçoivent ainsi les designs innovants avec une familiarité et une aisance cognitive améliorée dans la seconde étape, devenant moins spécifiques donc plus attractifs. Notons que la différence d'appréciation des modalités innovantes et typiques est plus prononcée pour les participants s'intéressant à l'art, ces participants prenant plus de temps pour regarder les designs les plus innovants et les notant mieux dès la première phase.

3.3.6.4 Complexité et attractivité

Hormis le caractère innovant, une autre propriété générale des objets influence l'évaluation : la complexité. La complexité d'un objet semble issue d'une part des propriétés des stimuli physique, tel que le nombre d'éléments circulaires, de couleurs ou de variétés de formes, et d'autre part des principes de groupement tel que la symétrie, ou encore de la typicalité [Locher01]. En 1970 Berlyne étudie les effets de la complexité sur l'appréciation [Berlyne70]. Il emploie des formes artificielles en faisant varier la complexité perceptive (par exemple stimulus avec de nombreuses textures) ou cognitive (stimulus avec des lignes représentant plusieurs symboles imbriqués). Il choisit d'associer la complexité avec le potentiel d'éveil d'un stimulus. Selon ses résultats, la préférence est liée au niveau de complexité du stimulus qui ne doit pas trop « éveiller l'attention » ni être « ennuyeux ». Berlyne [Berlyne71] propose alors un modèle en forme en U inversé pour définir la relation entre complexité (ou potentiel d'éveil d'un stimulus) et préférence. Des exceptions ont cependant été reportées [Martindale90] car le potentiel d'éveil apparaît comme subjectif, donc variant selon les individus et de leur capacités d'adaptation.

Carbon et collaborateurs utilisent ici aussi le paradigme « RET » pour mesurer l'évolution de l'appréciation d'un produit en fonction du critère de complexité [Leder05] sur l'étude d'habitacles de véhicules. Une première expérience montre une corrélation entre le nombre et angle de courbures

avec le niveau de complexité perçue, c'est-à-dire que plus l'habitable est composé de courbes, moins il apparaît complexe. Selon des études sur l'appréciation des visages [Etcoff99] et selon les principes de la gestalt [Arnheim54], la courbure semble induire des émotions plutôt positives car les formes douces et courbes sont plus souvent associées à la beauté et la douceur tandis que des traits et coins sont liés plutôt à des dimensions techniques et analytiques. Carbon observe ensuite que la note d'attractivité d'un habitacle est significativement affectée par le caractère complexe du stimulus, et que les modalités visuelles les plus complexes sont d'abord rejetées puis appréciées.

Ces différents travaux montrent que l'appréciation de produits peut être impactée par la manipulation de la typicalité des produits et de la complexité du stimulus [Langlois90], [Martindale88]. Tant le caractère innovant que la complexité perçue sont des facteurs qui influencent l'aisance de traitement, qui en retour produit une plus ou moins grande appréciation des stimuli. Plus le traitement est simple et rapide, plus l'aisance est forte, plus le prototype est apprécié, mais plus l'implication (c'est-à-dire la profondeur des traitements) est forte, plus la nouveauté est appréciée. On tire de cette section que les processus évaluatifs sont dépendants de réactions affectives et cognitives, issus de mécanismes non volontaires. Or, il apparaît que ces processus peuvent se manifester comportementalement par l'orientation du regard. C'est pourquoi la partie suivante présente le phénomène de « gaze cascading effect », consistant en un biais d'orientation du regard en faveur d'un élément en situation de choix multiple.

3.3.6.5 Le comportement d'orientation

Le comportement d'orientation est une réaction humaine fondamentale face à la nouveauté, qui est innée, automatique et protectrice [Pavlov27]. Ce comportement consiste en une priorisation des informations sensorielles pour traiter un stimulus nouveau et potentiellement dangereux. C'est ainsi que l'état affectif de surprise est associé à des changements cognitifs, émotionnels et hormonaux, produisant pendant quelques instants une réaction de paralysie temporaire [Sokolov63]. Le réflexe d'orientation est la forme la plus primitive du comportement d'orientation où son rôle premier est de préparer rapidement l'organisme au comportement de réaction « fight or flight », c'est-à-dire se battre ou s'enfuir. Le comportement d'orientation peut être involontaire (réflexe) ou volontaire et joue un rôle fondamental dans les processus perceptifs et cognitifs. Puisque la vision est la modalité perceptive dominante des humains, le regard suit fortement ces processus d'orientations somatiques, car il est le principal moyen de collecte d'informations sur un stimulus de l'environnement. C'est ainsi que le comportement d'orientation du regard est un vecteur de communication, car considéré comme une marque d'attention volontaire ou involontaire signalant l'intérêt, la préférence ou le désir de collaborer [Emery00]. Friesen et Kingston [Friesen03] montrent que nous sommes très sensibles à la direction du regard d'autrui, qui influence directement le temps de détection d'une cible ou l'attractivité d'un objet observé par un tiers. Fantz [Fantz64] montre qu'une longue inspection visuelle d'un objet par un enfant se corrèle avec une forte appréciation. Le regard est donc un moyen de collecte d'informations mais aussi un pointeur externe de l'état cognitif, intégré dans le processus de prise de décision et de l'expérience affective. Les mouvements des yeux ont donc non seulement une fonction d'intégration de l'information visuelle, mais aussi d'influence sur le choix et l'attribution affective.

L'observation de la formation de l'évaluation est difficile car l'introspection puis le report verbal ne sont pas nécessairement représentatifs des critères de décision. Les explications post-hoc du choix ne reflètent pas les critères peu clairs ou peu définis dans le choix, d'où l'intérêt d'observer le comportement du regard car manifeste l'orientation de l'attention. Simion ([Simion05], [Shimojo03]) mène un ensemble d'expérimentations démontrant que le regard assiste l'évaluation d'objets et la prise de décision. Un paradigme de comparaison par paire est employé pour comparer des expressions

faciales ainsi que des objets de consommation courant. Les consignes sont soit de choisir l'item que l'on préfère, l'item que l'on apprécie le moins, ou celui qui comporte le plus de formes circulaires. Les mouvements du regard sont enregistrés. Il apparaît une forte corrélation entre le choix et le temps d'inspection par produits, spécifiquement dans la dernière seconde précédant la décision. Dans une tâche de choix de préférence par paire, l'élément le plus observé dans les dernières secondes est le plus apprécié, et plus un élément est apprécié, plus il est observé. Il nomme ce biais de choix le « gaze cascading effect »¹³. La possibilité que le biais soit dû à une orientation du regard pour des raisons de mémorisation est éliminée par le fait qu'après la décision, le regard se redistribue de manière équitable sur les deux modalités visuelles. Une deuxième expérimentation montre que le « gaze cascading effect » est lié au « mere exposure effect ». Le dispositif consiste à afficher les photos de deux produits à comparer de la manière suivante : le produit A est affiché à gauche de l'écran pendant 300ms, puis le produit B est affiché à droite pendant 900ms. Lorsque cette séquence est répétée plus de 6 fois, il apparaît que le produit le plus longtemps affiché est significativement préféré. Ainsi, plus une image est affichée longuement, plus elle est appréciée. Le regard est donc plausiblement influencé par la préférence, mais la préférence est aussi imputable à l'orientation du regard. De manière intéressante, avec le même dispositif, mais dans une tâche contrôle où les images au lieu d'être affichées alternativement à droite et à gauche de l'écran le sont au centre, le biais de choix disparaît. Tout se passe comme si le traitement du flux visuel était inhibé lorsque le regard ne doit pas se réorienter, ou comme si la orientation somatique était nécessaire à l'apparition du biais. Notons enfin que ces expériences montrent que l'évaluation de l'élément le moins apprécié semble plus similaire à l'évaluation de l'élément comportant le plus de courbes que l'évaluation de la préférence, en termes d'orientation du regard et de manifestation du « gaze cascading effect ». Ainsi, la décision de rejet pourrait être basée sur des critères plutôt objectifs, à la différence de la décision de préférence s'appuyant sur des critères subjectifs. En effet, la tâche de rejet ne produit pas le même renforcement perceptuel par augmentation de l'observation de l'item choisie au contraire de la tâche de préférence.

Cette section aura permis de faire le lien entre le comportement oculaire et la prise de décision, où le regard est non pas un simple capteur mais un acteur important de l'élaboration du jugement.

3.3.7 Conclusion

Ce chapitre portant sur les processus évaluatifs montre que l'évaluation d'objets se base sur le traitement d'informations provenant d'attributs visuels et que la relation entre la décision et les informations perceptives acquises n'est pas directe car le jugement peut être élaboré à différents niveaux et sous l'influence de nombreux biais. Cependant, le regard est non seulement un pointeur externe de l'activité cognitive de prise de décision (« gaze cascading effect ») mais aussi un acteur de cette décision. La formation d'impressions sur des objets manufacturés est donc l'aboutissement de processus nécessitant l'acquisition d'informations locales (analyse attentionnelle détaillée) et aussi globales (identification, catégorisation). La complexité des mécanismes orientant l'attention nécessite d'adopter une approche statistique des mesures oculométriques afin d'extraire les liens entre les processus oculomoteurs et le traitement de l'information visuelle en situation d'évaluation visuelle. Ainsi, après avoir présenté dans cette partie théorique les propriétés des scènes visuelles que l'on souhaite exploiter, puis les différents facteurs qui guident le regard et enfin les processus de traitement de l'information visuelle associés à l'évaluation, nous nous intéressons à la mesure

¹³ « gaze cascading effect » ou effet cascade du regard est dénommé ainsi car au cours du choix se manifeste un effet en cascade : plus on regarde un élément, plus on l'apprécie, et plus on apprécie un élément, plus on le regarde.

oculométrie proprement dite. La partie suivante expose ainsi différentes méthodes d'analyse des données oculométriques ayant pour but de caractériser et catégoriser les trajets oculomoteurs. On va s'attacher plus particulièrement à mettre en évidence les informations visuelles spécifiquement traitées selon les conditions expérimentales. Ces méthodes seront ensuite directement appliquées dans notre cadre expérimental.

- Volet 2 -
PARTIE EXPERIMENTALE

Chapitre 4

Cadre expérimental

4.1 Présentation

Les éléments théoriques que nous avons abordés précédemment montrent que la perception visuelle des scènes réelles sollicitent de nombreux processus (de détection, d'identification, de catégorisation,...), traitant en parallèle de grandes quantités d'informations visuelles, activant des mécanismes d'interactions permanentes entre l'orientation comportementale du regard et le système cognitif. De plus, l'évaluation d'un objet est influencée par les propriétés intrinsèques mais aussi par des facteurs contextuels telles que les attentes du participant, les pré-connaissances, les interférences entre les différentes modalités sensorielles, aboutissant à de nombreux biais de jugement. Cependant le comportement du regard semble être un processus manifestant l'orientation de l'attention visuelle mais aussi l'intérêt et l'attractivité. On peut donc poser l'hypothèse que malgré la complexité des processus impliqués la mesure des mouvements du regard lors de tâches d'évaluation d'objets manufacturés devrait fournir des indications quant aux zones visuelles ayant attiré l'attention et ainsi lier les processus d'attention visuelle et d'évaluation. En effet, si l'on peut inférer à partir de l'orientation de l'attention visuelle les processus activés lors de l'évaluation de produits manufacturés, cela permet d'optimiser le développement du produit en phase de conception pour l'adapter au mieux aux intentions ciblées. Des études menées précédemment confirment l'intérêt d'explorer ce champ méthodologique : Hammer et Lengyel [Hammer89] montrent que les participants observent les zones de design de l'objet les plus informatives pour la reconnaissance de la marque ; l'attention s'est donc orientée vers des attributs visuels particuliers, et les informations visuelles traitées ont pu être inférées à partir du regard. Ce même auteur [Hammer92] dans une expérience ultérieure, montre que le regard se pose sur des zones spécifiques à la consigne. Différentes études sont menées par des équipes de recherche finlandaises [Sharmin04] sur l'utilisation de mesures oculométriques pour l'aide à la conception d'objets manufacturés, montrant notamment que le niveau d'expertise influence la manière d'observer le produit, et donc les attributs visuels traités pour l'évaluation. Des travaux effectués sur les processus de catégorisation d'objets manufacturés [Dompnier06] mettent en évidence l'existence de réseaux de propriétés, liant les attributs d'un produit à son évaluation par l'apparition d'inférences. Ces attributs, visuels dans notre cas, sont donc support de l'attention et intégrés pour l'évaluation. On peut donc émettre l'hypothèse que traiter les données oculométriques peut permettre d'une part de mieux comprendre les différents processus attentionnels activés dans une tâche évaluative, et d'autre part d'inférer, à partir des comportements du regard, les informations visuelles traitées pour le jugement.

Notre travail repose sur plusieurs niveaux de problématiques. Le premier niveau est celui du traitement des données. D'un point de vue analyse statistique du signal, les mesures oculométriques peuvent être considérées comme un ensemble de données nécessitant plusieurs phases de traitements afin d'établir des descripteurs (caractériser les séquences oculomotrices, et de faire émerger des composantes communes (catégoriser ces séquences)). Cependant, parce que ces données sont des mesures du comportement humain, le deuxième niveau de notre travail va porter sur les processus perceptifs et cognitifs impliqués dans l'évaluation visuelle d'objets manufacturés. On

cherchera ainsi à limiter les effets des variabilités idiosyncrasiques afin de faire émerger les facteurs d'orientation de l'attention visuelle, et les informations visuelles spécifiquement traitées.

Cette partie est composée d'un chapitre portant sur les méthodes d'analyse et de traitement des données, puis de quatre phases expérimentales. La première, composée d'un paradigme d'évaluation monadique et d'un paradigme d'évaluation par paire, a pour objectif d'observer les effets de la tâche évaluative et des propriétés des modalités visuelles sur le comportement du regard. Différentes modalités visuelles sont ainsi évaluées par plusieurs participants. Le paradigme d'évaluation monadique, consistant en une évaluation d'un seul produit exposé, est la situation la plus courante d'évaluation de produits. Le paradigme d'évaluation par paire, limitant le biais d'observation central de l'image, doit inciter une évaluation basée sur un processus analytique de comparaison de critères visuels plutôt qu'un processus d'évaluation heuristique sur la base d'informations mnésiques ou affectives. Il doit donc favoriser la sélection par l'attention de zones visuelles locales. Pour ces expérimentations, on appliquera différentes analyses spatiales et spatio-temporelles du comportement du regard, afin de vérifier :

- si les attributs physiques (texture, luminance, formes,...) ou les propriétés sémantiques (liste d'objets présents dans la scène et les attributs de jugement associés -qualité, évocation, cohérence) participent à l'orientation du regard et à l'élaboration du jugement
- si les valeurs sémantiques que le concepteur a attribué au produit sont identifiées par les participants en se basant sur des éléments visuels spécifiques
- s'il existe des classes de stratégies exploratoires en fonction des attributs du produit vers lesquels l'attention se dirige.

La seconde phase de cette partie expérimentale est complémentaire à la précédente afin de vérifier des hypothèses spécifiques au traitement des informations visuelles. Cette étude, qualitative, consiste en un paradigme d'évaluation d'objets en situation monadique et par paire lorsque le regard est contraint au centre de la scène. Ce travail doit permettre d'observer si un pré-choix est possible sans déplacer le regard, donc si le traitement des informations visuelles pertinentes pour la tâche nécessite des fixations locales ou si les informations globales sont suffisantes.

La troisième phase propose de mettre en évidence le rôle de pointeur du regard. En effet, l'orientation de l'attention a non seulement pour fonction de contrôler la capture des informations visuelles, mais aussi de manifester une attitude interne exprimant l'intérêt ou l'attractivité. De plus, il a été observé que l'orientation du regard biaise le jugement, par l'action somatique en elle-même. On va ainsi chercher à mettre en évidence l'interaction entre l'orientation du regard et les produits évalués ('gaze cascading effect') [Simion05]. Pour cela, on étudiera le comportement oculomoteur de participants évaluant les habitacles sous des consignes de préférence et de catégorisation avec une durée d'exposition libre et non pas contraint comme dans la phase précédente.

La quatrième phase va présenter plusieurs méthodes d'analyse des processus attentionnels afin de mettre en évidence les facteurs guidant le regard. La première partie présente une méthode d'estimation des zones visuelles de décision, nommée paradigme des « Bubbles » [Gosselin01] consistant à dégager les relations statistiques entre la visualisation d'une zone locale et la décision. La seconde partie présente une analyse des effets de la saillance visuelle sur le guidage du regard. Enfin la troisième partie propose un modèle d'estimation de la contribution de différentes hypothèses de guidage du regard basé sur un algorithme de mélange de gaussiennes, pour les situations d'évaluation monadique et par paire, avec et sans consigne.

Cadre expérimental général	Chapitre 4	Introduction de la problématique Spécificité des paradigmes
Méthodes de traitement des données	Chapitre 5	Méthodes d'analyse des données oculométriques
Partie expérimentale 1	Chapitre 6	Evaluation monadique
	Chapitre 7	Evaluation par paire
Partie expérimentale 2	Chapitre 8	Evaluation en situation de vision contrainte
Partie expérimentale 3	Chapitre 9	Comportement d'orientation du regard et « gaze cascaded effect »
Partie expérimentale 4	Chapitre 10	Evaluation via le paradigme des « Bubbles » Exploration libre et saillance Modèle d'évaluation des contributions des facteurs attentionnels à la distribution des fixations oculaires
Synthèse et conclusion		

La section qui va suivre expose plus spécifiquement les différents matériels expérimentaux qui vont être employés dans les études. Nous aborderons ensuite « la boîte à outils » de méthodes d'analyse et de traitement des données oculométriques qui seront appliquées.

4.2 Spécificités des paradigmes employés

✓ *Spécificité des scènes visuelles*

Notre travail s'intéresse à la perception et l'évaluation d'objets manufacturés photographiés. Ces scènes visuelles sont concrètes, c'est-à-dire sont cohérentes sémantiquement tant à un niveau global que local. Elles ne violent pas les lois de l'organisation perceptuelle naturelle. Cela implique que le regard ne va pas obéir à un processus de reconstruction de la scène. Celle-ci sera immédiatement appréhendable. Ensuite, ces scènes sont dépeintes (photographies), ce qui signifie que ce ne sont pas directement les objets qui sont traités, mais leur représentation bidimensionnelle ; cela implique que les informations de volume sont manquantes, pouvant alors perturber la prise d'information par l'apparition d'un processus de recherche d'indices liés à la profondeur, et produisant des lacunes d'informations pour la résolution de la consigne. Enfin le traitement visuel sera simultané sur plusieurs plans de l'espace, limitant la possibilité de distinguer différents objets selon leur plan.

✓ *Particularité des objets*

Le produit automobile, et encore plus l'habitacle, est un objet particulier. Tout d'abord il s'agit d'un objet à part entière, dénommé donc catégorisé en tant qu'entité propre. Cependant, il s'agit aussi d'une scène, car cet objet super-ordinal est composé d'un ensemble de sous-objets qui peuvent être reconnus et dénommés individuellement (volant, compteur, siège...). On peut donc considérer l'habitacle soit comme un objet global qui peut soit être appréhendé sans décomposition en sous-entités, soit être considéré comme une scène d'objets qui attireront individuellement plus ou moins

l'attention. L'existence de ces différents niveaux sémantiques complexifie l'identification des éléments attirant l'attention.

Une seconde particularité de la scène étudiée provient de la densité spatiale des objets qui la composent. En effet, l'habitacle automobile est composé d'un ensemble d'éléments superposés ou accolés. Même si le système perceptif est capable d'appréhender individuellement ces éléments, la largeur du champ visuel et la densité d'objets dans l'environnement immédiat, quelle que soit la position du regard dans la scène, ne permettent pas, le cas échéant, de définir l'objet local vers lequel est dirigé l'attention visuelle.

Une troisième particularité est la forte pré-connaissance culturelle et sociale du produit automobile. Une très grande partie des européens possède une voiture, et il s'agit désormais d'un produit de consommation courante. Il existe de ce fait chez tous les participants une forte pré-connaissance de l'objet étudié, au moins en termes d'organisation spatiale des objets dans la scène et de structure (liste de sous-objets composant nécessairement l'habitacle). Cette pré-connaissance est encore plus forte chez les personnes ayant un intérêt particulier pour ce produit, car possédant une connaissance approfondie de l'objet leur permettant de l'évaluer rapidement sur la base de critères préétablis. Le déploiement de l'attention sur la scène est influencé par cette forte pré-connaissance, car la reconnaissance est optimisée par des facteurs mnésiques, et parce que la représentation cognitive de la scène sera soumise à des inférences : la présence d'un objet implique la présence d'un second, qui sera considéré comme présent sans avoir été perçu. Ainsi l'interprétation de l'orientation de l'attention doit se faire avec précaution car la pré-connaissance va influencer le traitement des informations visuelles.

Le choix des objets étudiés est motivé par la nécessité de devoir comparer les comportements visuels et les jugements entre ces objets. Ils doivent donc être suffisamment similaires en termes d'organisation spatiale, mais suffisamment différents en terme visuel et/ou sémantique pour permettre une différence d'évaluation. Des habitacles de véhicules doivent ainsi être sélectionnés selon ces critères. Le choix des produits (habitacles de véhicule) s'est effectué tout d'abord sur proposition du partenaire industriel qui souhaitait avoir des résultats sur quelques produits spécifiques, et par la volonté de comparer des produits (a) qui sont, en terme de gamme, connus (ou utilisés) par le grand public, et (b) qui sont haut de gamme donc moins bien connus. C'est ainsi que le produit 207 (gamme de base) et le produit C6 (haut de gamme) ont été choisis. Pour le choix des typages (modalités visuelles d'un même produit), pour 207 a été choisie la 207-sport du fait de la présence de signes sémantiques locaux exprimant la sportivité ; la 207-standard a été choisie en tant que référence. Pour C6, seules deux déclinaisons existent : blanche et noire, ce qui convenait car présentant peu de différences locales mais globales. Les deux types de différences entre modalités (informations locales *versus* globales) étaient ainsi représentés.

✓ *Nature des consignes*

L'un des objectifs de notre travail est de relier l'orientation de l'attention visuelle et la recherche d'informations propres à la consigne, dans le but de confronter le processus d'évaluation du produit et les zones visuelles traitées. C'est pourquoi on s'intéresse aux apports de l'oculométrie dans l'évaluation du produit automobile. Les stimuli sont des habitacles de véhicules, les consignes doivent donc être pertinentes à appliquer sur des photographies d'habitacles et elles doivent nécessiter des mouvements du regard. Ces consignes ne doivent pas faire appel à une forte expertise car les participants sont des individus non experts du monde automobile. Puisque l'on souhaite comparer les comportements oculomoteurs dans différentes situations, plusieurs consignes sont définies. Une première consigne doit faire appel à des éléments visuels locaux tandis que la seconde peut être plus globale. L'une des consignes doit être orientée de telle sorte qu'elle vérifie une intention des concepteurs et une autre consigne peut être plus subjective.

C'est ainsi que la consigne « sportivité » est sélectionnée en tant que consigne basée sur des éléments « locaux », avec une intention sémantique concepteur, dans une catégorie facilement définissable et consensuellement reconnue. La consigne « niveau de gamme » est choisie en tant que consigne basée sur des informations globales mais, selon les constructeurs, discriminent les deux modalités visuelles de 207. La consigne « caractère raffiné » est choisie en tant que consigne évaluative globale et sans hypothèses constructeur. On peut remarquer que la nature de ces consignes est très différente de celle d'une consigne de recherche de cible ou de catégorisation rapide par exemple. En effet, nos consignes relèvent de jugements faisant appel à la fois à une pré-connaissance de signes visuels locaux exprimant une propriété (présence de chromes par exemple), à la détection de ces signes, et à l'application d'heuristiques de décision.

✓ *Types de recueils*

Cherchant à relier le comportement du regard, le traitement de l'information visuelle et l'évaluation d'objets, il est nécessaire de mettre en place une procédure de recueil de jugement par le participant. Les paramètres suivants sont à définir pour chacun des paradigmes :

- durée d'exposition : le participant est-il libre de mettre fin librement à la tâche ou s'agit-il d'une durée d'exposition fixe ? On cherche dans la première partie expérimentale à relever les éléments du comportement visuel communs à tous les participants pour une même condition expérimentale. Il est donc pertinent de fixer la durée d'exposition afin de pouvoir regrouper l'ensemble des essais et de comparer les différentes conditions expérimentales. Dans la seconde partie expérimentale, l'un des objectifs est d'observer le comportement du regard au moment de la prise de décision ; il est donc nécessaire de laisser le participant mettre fin lui-même à l'essai afin d'identifier le moment où il considère avoir terminé la tâche.

- recueil de l'évaluation : doit-il se faire post-exposition ou non (« offline » ou « online ») ? Il est intéressant de recueillir des indices quant à l'élaboration de la décision au cours de l'exposition, car il n'est pas démontré qu'il y ait un lien direct entre la phase de prise d'information et la phase de prise de décision (en termes d'orientation de l'attention). Cependant, hormis la procédure de pensée à voix haute (« think aloud »), il n'existe pas de moyens permettant de suivre la construction de cette évaluation. Cette procédure, dont les résultats sont complexes à analyser, risque d'interférer avec la consigne principale. C'est pourquoi il est choisi de recueillir l'évaluation après exposition (« offline »), en étant prudent dans l'interprétation des éléments verbalisés comme critères de décision.

- évaluation : doit-elle être globale, en fournissant une note, ou bien doit-elle être justifiée par des attributs décisionnels locaux ? Puisque l'on cherche à relier la prise d'information locale et les critères de décision, il est pertinent de demander aux participants sur quels critères leur décision a été fondée. Cependant, la nature de la consigne ne permet pas nécessairement de justifier le choix (lorsque le choix est subjectif ou lorsque les critères de décision sont ambiants et non locaux). Ainsi, la première réponse fournie par le participant est une note pour l'évaluation monadique (indiquant un niveau global d'évaluation) ; pour la comparaison par paire la première réponse est le choix, puis la note de chaque élément. Les notes sont fournies par le déplacement d'un curseur sur une échelle en 9 pas, de pas du tout [...] à très [...]. Dans les deux cas il est ensuite demandé au participant de justifier son choix par la verbalisation de critères de décision. Il est nécessaire de traiter ces données avec précaution tant les processus qui interviennent, biaisent ou interfèrent avec l'évaluation sont complexes et méconnus.

✓ *Participants*

Chaque condition expérimentale est réalisée par au moins 10 participants. Ces participants sont recrutés de manière à obtenir une répartition homogène des âges (de 20 à 65 ans) et des sexes. Ils doivent avoir acheté une voiture neuve du segment B (bas) dans les deux dernières années, ne pas

travailler dans le domaine de l'automobile ou du design, et ne pas avoir fait de test clientèle dans l'année en cours. Les participants sont rémunérés 40 euros pour le test.

✓ *dispositif*

L'évaluation se fait dans une salle obscure, sur un écran de vidéo-projection à une résolution de 1400*1050, soit 125*167cm, à une distance de 2 mètres (40°x34° du champ visuel). Les mouvements oculaires sont enregistrés avec l'oculomètre Facelab® 4.1, un système déporté, réglé en mode précision (environ 1° de précision) et utilisation des émetteurs infrarouges.

✓ *Notes sur les termes lexicaux*

Dans ce travail expérimental, nous manipulons différents stimuli visuels sous plusieurs consignes. Plusieurs termes doivent ainsi être précisés.

- « Essai » : ensemble des recueils (oculomoteurs, notes ou verbalisations) lors d'une évaluation
- « condition expérimentale » : désigne un ensemble d'essais d'évaluation avec une même consigne et sur un même stimulus,
- « produit » : dans les expérimentations 1 et 2, nous étudions deux produits, 207 et C6
- « modalité visuelle » : désigne un stimulus ; par exemple il existe deux modalités visuelles de 207 : 207-Sport et 207-Standard. Les termes « versions d'habitacle » ou « typages » y font aussi allusion,
- « Figure » : Principalement utilisé lors de l'exposition par paire, ce terme est similaire à celui de « modalité visuelle » mais cible spécifiquement l'un des deux éléments de la paire.

✓ *Synthèse*

L'apport de l'oculométrie à la compréhension des processus d'évaluation des objets manufacturés peut être intéressant dans l'objectif d'identifier les attributs du produit ayant attiré l'attention et influencé la décision, mais aussi pour mettre en évidence les différents processus de la perception, de l'identification, et de la prise de décision activés dans ce type de tâche. Les expérimentations doivent fournir un nombre suffisant de mesures afin de pouvoir traiter les données par des outils statistiques, de limiter les variabilités idiosyncrasiques et mettre en évidence les facteurs les plus robustes en fonction des conditions expérimentales. Les paradigmes employés ont pour particularité de solliciter un traitement de l'information visuelle fortement parallèle du fait de la densité spatiale des éléments visuels et de la nature non ciblée des critères de résolution de la consigne. De plus, les différents niveaux de catégorisation sémantique des objets de scène complexifient le lien entre les processus de prise d'information et les processus de décision. Ces éléments impliquent une faible répétabilité du comportement, donc une forte variabilité interindividuelle et intra-individuelle. Les interactions entre les modalités visuelles, la consigne, et les spécificités cognitives du participant font de chaque essai un cas unique. L'un des principaux objectifs est donc d'appliquer des méthodes d'analyses des mesures oculométriques faisant émerger des informations suffisamment stables pour permettre une interprétation du comportement oculomoteur. Les différents types de traitements de données oculométriques sont présentés dans le chapitre suivant.

Chapitre 5 Traitement des données oculométriques

5.1 Acquisition des données et pré traitements

Traiter statistiquement les données oculométriques consiste à mettre en lien les variables oculomotrices avec les conditions expérimentales, à observer par des représentations multidimensionnelles les facteurs guidant l'attention visuelle, puis à les modéliser. L'objectif de cette section n'est pas de présenter un ensemble exhaustif d'hypothèses de traitement cognitif des informations visuelles et leur traitements de données oculométriques associés, mais plutôt d'exposer les principales familles d'outils statistiques à notre disposition ainsi que les procédures d'analyses typiquement employées dans le cadre de nos paradigmes expérimentaux. Les traitements spécifiques seront présentés lors de leur utilisation. On notera que les traitements présentés ici correspondent à une analyse *a posteriori* des mesures, c'est-à-dire que les outils d'analyse en temps réel du signal fourni par « l'eye-tracker » ne seront pas présentés.

Mesurer le comportement oculomoteur consiste à relever les coordonnées de la projection de la direction du regard sur l'écran. Ce point, défini en $\{x, y\}$ et noté \vec{x} correspond au centre de la fovéa dont le diamètre varie entre 3 et 5 degrés. x est la coordonnée en pixel sur l'axe des abscisses dans le repère de l'écran, et y est la coordonnée sur l'axe des ordonnées. La valeur de la profondeur z de l'objet regardé n'est pas relevée étant donné que l'on travaille sur écran. La mesure est dépendante de l'échantillonnage du système, qui prélèvera à une certaine fréquence la position du regard. Cette mesure échantillonnée est nommée impact. La période d'échantillonnage dépend du type d'oculomètre et de leur paramétrage. Ainsi, le Facelab 4.1 de Seeing Machine® fonctionne à 60Hz, c'est-à-dire une mesure d'impact toute les 16,7ms. L'eyelink2® fonctionne à 250hz avec la mesure des deux yeux et à 500hz avec un seul œil mesuré, soit respectivement 4 et 2 ms de délai entre les impacts.

Le signal mesuré peut être ponctuellement corrompu, par exemple lors de clignements d'œil ou de détournements de tête. En appliquant une interpolation linéaire géométrique des deux mesures encadrant la perte du signal, il est possible de reconstituer une tendance de trajectoire du regard. Considérant qu'un clignement d'œil est d'environ 80ms et qu'une fixation dure au moins 120ms, il est possible de négliger la perte de signal dû à cet événement, et une interpolation linéaire sera une approximation suffisante.

On s'intéresse au comportement balistique du regard, d'où la nécessité de transformer la séquence des impacts mesurées en séquences de fixations et saccade. Différentes approches sont proposées pour détecter une saccade et de manière duale un point de fixation, à partir des points d'impacts échantillonnés :

- Détection basée sur la vitesse de déplacement du regard
- Détection basée sur l'accélération du déplacement
- Détection basée sur la dispersion spatiale des points de mesure

Ces approches peuvent être utilisées isolément ou de manière combinée, avec des algorithmes simples de seuillage ou plus complexes, notamment les chaînes de Markov caché (système d'apprentissage reconnaissant les motifs) ([Salvucci00], [Salvucci01]). Cependant, les systèmes oculométriques du marché n'utilisent que les algorithmes de seuillage, du fait de leur simplicité de mise en œuvre et de leur robustesse. Selon le type d'applications, donc des processus attentionnels observés, la détection des mouvements oculaires nécessitera des réglages différents. Ainsi, le système EyeLink2® utilise dans son mode « utilisation cognitive » des paramètres permettant une bonne stabilité du signal, une faible sensibilité au bruit, et ignore les micro-saccades de moins de $0,6^\circ$; le seuil de vitesse est fixé à $30^\circ/s$, le seuil d'accélération à $8000^\circ/s^2$, le déplacement angulaire de $0,6^\circ$. Alors que le mode « psychophysique » paramétré à $22^\circ/S$ pour la vitesse, $4000^\circ/s^2$ pour l'accélération, sans limite inférieure de déplacement angulaire, permet de relever les plus faibles saccades et micro-saccades, mais ne permet pas une aussi bonne stabilité du regard que le mode cognitif. Pour le système GazeTracker®, le seuil de vitesse est de $40^\circ/s$, l'accélération à $9000^\circ/s^2$, une dérive angulaire supérieure à $0,1^\circ$, et une durée minimale de saccade de 40ms. Plus le seuil de détection du mouvement sera fin et plus la fréquence de mesure sera grande, plus des processus fins et rapides pourront être mesurés, mais plus les données seront lourdes à traiter. Nous utilisons un algorithme multicritère de détection des fixations par seuillage identique au mode « cognitif » de EyeLink2 car les processus que nous souhaitons observer sont de l'ordre de la seconde et non pas de la milliseconde. De plus, ces algorithmes de détection multicritères sont stables et très simples à implémenter. Sur la base de notre expertise, nous rajoutons une condition de filtrage du signal à la détection des saccades : la durée d'une fixation doit être au moins égale à 80ms.

A partir de la liste $\{t,x,y\}$ des impacts, les fixations sont identifiées, et les caractéristiques suivantes sont établies :

- Rang de la fixation dans la séquence
- Position de la fixation (centre de gravité des impacts associés)
- « Timecode¹⁴ » de début de fixation : « timecode » du 1^{er} impact
- « Timecode » de fin de fixation : « timecode » du dernier impact
- Durées des fixations
- Amplitudes, vitesses, accélérations des saccades

L'association entre la position de la fixation et l'information visuelle traitée doit toujours se faire en considérant que le point mesuré n'est que le centre du flux visuel et pas l'unique zone depuis laquelle l'information visuelle est extraite. Non seulement le traitement de l'information visuelle se fait sur l'ensemble du champ visuel ($120-160^\circ$), mais le diamètre de fovéa (zone de l'espace dans laquelle la qualité du champ visuel est maximal) est de l'ordre de 3 à 5° . Les coordonnées fournies par la mesure est donc le centre géométrique d'un disque dont le rayon peut varier entre 1,5 et 2,5 degrés selon la nature du stimulus et de la tâche. On devra en tenir compte dans la précision spatiale des interprétations sémantiques de la prise d'information visuelle, mais aussi dans le paramétrage de nos algorithmes.

Les analyses exploratoires préliminaires pour caractériser les relevés de données oculométriques concernent principalement des statistiques descriptives mono-variées (moyenne, écart-type, histogramme) des durées de fixation, des amplitudes de saccades, des vitesses de saccades..., et

¹⁴ Le Timecode correspond à la durée en seconde ou millisecondes écoulée depuis le début de l'exposition. Il s'agit donc d'un temps propre à la séquence de mesure.

également des observations conjointes telle que la durée de fixation en fonction de l'amplitude de la saccade précédente ou du rang [Velichkovsky05]. A ces descriptions globales peut se rajouter une description spatiale « du chemin visuel » parcouru en calculant la surface couverte (cumul de zones fovéales fixées) ou le chemin parcouru (cumul des amplitudes de saccades). Cette liste d'indices n'est pas exhaustive, ce sont des indices globaux tant sur la dimension temporelle que spatiale. Dans la suite, nous allons développer les techniques permettant des analyses locales temporellement et spatialement.

En synthèse, ces variables descriptives permettent de caractériser et catégoriser les différentes séquences oculomotrices, et de tester l'existence de stratégies oculaires exploratoires par individus. Cependant, ces données descriptives ne sont représentatives que des séquences individuelles. Parce que les facteurs de guidage du regard interagissent et sont instables, mais aussi parce qu'ils s'enchaînent dans le temps, ces mesures par individus sont soumises à un fort bruit. De plus, ces variables ne permettent pas de comparer les composantes oculomotrices statistiquement représentatives de chaque condition expérimentale. C'est pourquoi nous nous intéressons aux méthodes d'estimation des densités spatiales des fixations et aux méthodes multi variées des analyses des zones d'intérêt.

5.2 Estimation et exploitation des distributions de fixations : densités et classifications

Dans le but d'obtenir des informations sur le comportement du regard, une première étape consiste en une analyse des fonctions de distribution de densité des fixations sur la scène. Les cartes de densité sont particulièrement informatives car elles représentent la distribution spatiotemporelle de la prise d'information visuelle sur la scène. De plus, ces fonctions de distribution peuvent être comparées par des indices statistiques puis rapportées aux conditions expérimentales. Les cartes spatiales et spatio-temporelles de densité permettent entre autres de diagnostiquer le comportement du regard sur l'image, si celui-ci se focalise sur des zones spatiales précises ou s'il se diffuse plus uniformément sur l'image. Dans le premier cas, il sera alors pertinent de poursuivre l'étude par la caractérisation des zones d'intérêt, alors que dans le second cette analyse n'aura pas lieu d'être.

Une seconde étape d'analyse du comportement oculomoteur consiste à appliquer des méthodes de classification des points de fixations afin de créer des groupes, qui seront considérés comme autant de zones d'intérêts.

Ces deux types de traitement (densité et classification) ont l'intérêt de pouvoir être appliqués sans hypothèse préalable sur le comportement du regard.

Deux méthodes d'estimation des densités de fixations sont proposées : méthode d'estimation par noyaux de Parzen, et méthode d'estimation par Mélange de Gaussiennes. Plusieurs outils permettant de comparer ces modèles de densité sont ensuite exposés. Enfin, les méthodes de classification des points de fixation que nous emploierons sont présentées.

5.2.1 Estimation des fonctions de densité : distribution spatiale des fixations du regard sur la scène

Deux méthodes complémentaires d'estimation de densités seront décrites. La première (Noyaux de Parzen) est une technique non paramétrique qui est très rapide en temps de calcul et qui permet une analyse temporelle en effectuant l'estimation sur des fenêtres temporelles glissantes. La seconde

(Mélange de Gaussiennes) est une technique paramétrique qui est plus lourde en temps de calcul, mais qui permet d'avoir une équation analytique décrivant la distribution (utile pour les traitements ultérieurs de comparaison). Elle permet également de pouvoir réaliser une classification de points de fixation sur les zones d'intérêt.

Tant pour la méthode d'estimation des densités par noyaux de Parzen que pour la méthode d'estimation par mélange de gaussiennes, les données d'entrée sont les points de fixation de l'ensemble des essais¹⁵ effectués pour une même condition expérimentale (i.e. même image, même consigne) quel que soit le participant. La variabilité interindividuelle est alors limitée par l'exploitation d'un nombre de points de fixation estimé statistiquement suffisant. Les fixations oculaires sont caractérisées par 4 valeurs, à savoir les coordonnées spatiales (x,y) de la position du point de fixation sur l'écran, la durée de la fixation en secondes et le timecode (délai en secondes depuis le début de l'exposition) de début de fixation.

L'utilisation de la variable « durée de fixation » est possible mais pas indispensable. Elle est fonction de la signification cognitive qu'on leur attribue. Si l'on considère que la durée de la fixation n'est pas un indicateur de la force de l'attention déployée localement, alors cette variable doit être ignorée et les densités sont établies uniquement sur la base des coordonnées spatiales. Si au contraire on considère que plus une zone est longuement fixée plus sa probabilité d'avoir attiré l'attention est élevée, alors la variable durée de fixation sera utilisée pour pondérer la contribution de la fixation au modèle de densité. L'emploi de cette pondération est paramétrable pour les deux algorithmes d'estimation des densités. Cependant, du fait de l'absence dans la littérature actuelle de preuves supportant l'idée que la durée des fixations soit un indice de l'activité de cognition visuelle dans la perception des scènes réelles, nous n'intégrerons pas cet indice dans les modèles de densité (du fait de l'impossibilité d'affirmer qu'une durée de fixation plus grande correspond à un intérêt local plus fort).

5.2.1.1 Méthode d'estimation de densité par noyaux de Parzen

Notons que pour la suite de l'exposé, le vecteur coordonnées {x,y} sera représenté par \vec{x} . L'estimation de la densité des fixations par la méthode des noyaux de Parzen consiste à positionner sur chaque lieu de fixation une fonction gaussienne de largeur spécifiée [Parzen62]. Toutes les fonctions sont ensuite sommées et l'espace normé. On obtient alors en chaque point de l'espace la probabilité que le lieu ait été fixé. Un paramètre qui reste à définir est l'écart-type σ de la fonction gaussienne. Si on considère un noyau gaussien de matrice de variance-covariance $\Sigma = \alpha \cdot \text{Id}$ (Matrice identité), alors la formulation est

$$f(\vec{x}) = \sum_{i=1}^N p_i \cdot N(\vec{x}; \vec{\mu} = \vec{x}_i, \sigma)$$

- $f(\vec{x})$ est la valeur de la densité estimée au point de coordonnées,
- \vec{x}_i est le vecteur coordonnées du ième point de fixation,
- $\vec{\mu}$ est le paramètre de centre de gaussienne,
- σ est le paramètre de largeur de gaussienne,

¹⁵ Un essai étant la séquence de fixations oculaires d'un participant, mesurée pour une condition expérimentale

- p_i est la pondération de la fonction gaussienne du i -ème point, qui permettra d'appliquer une pondération proportionnelle à la durée de fixation si on souhaite l'intégrer.

Le paramètre σ est fixé en prenant en compte les propriétés psychophysiques du système visuel. Puisque l'on applique une fonction gaussienne au centre de chaque point de fixation, on définira que la variance de cette fonction correspond à la taille de la fovéa projetée sur l'écran.

Cette méthode a pour avantage d'être rapide et déterministe. Elle permet de fournir des résultats visuels rapidement interprétables sur le déploiement de l'attention sur la scène. Elle est en outre plus spécialement adaptée à l'observation de l'évolution temporelle de l'attention. Etant une méthode non paramétrique, il n'y a pas une formulation analytique des modes de la distribution (zones de forte densité). En conséquence, cette méthode ne permet pas d'effectuer une classification des points de fixation aux modes. Ainsi, pour aller plus loin dans l'analyse des fonctions de distribution des fixations, la méthode de mélange de gaussiennes est alors utilisée.

5.2.1.2 Discrimination des modes de Parzen : algorithme « Meanshift »

L'algorithme « Meanshift » ([Fukunaga75], [Cheng95]) consiste à retrouver les modes d'une distribution de points, c'est-à-dire les zones les plus denses. L'algorithme est basé sur une estimation non paramétrique de la densité par une méthode des noyaux de Parzen, suivie d'un calcul local du gradient de densité. Les modes extraits correspondent aux maxima locaux de la densité. Comme pour la méthode des noyaux de Parzen, cette méthode est uniquement paramétrée par la taille de noyaux, définissant l'étendue spatiale de ces derniers. Ce paramètre spatial permet le contrôle du nombre de modes et donc de la classification à suivre : plus il est grand et moins le nombre de modes est élevée. C'est une méthode qui est très utilisée en analyse d'image (segmentation d'objets).

L'emploi de cette méthode est très aisé, d'une part du fait de sa souplesse d'utilisation et de sa rapidité d'exécution, mais aussi de part l'absence de paramétrages en fonction d'hypothèses *a priori*. Une limitation cependant est due au réglage du paramètre spatial : une valeur trop grande ne permettra pas de dégager des modes associés à des zones locales d'intérêt, et une valeur trop petite fournira un très grand nombre de modes avec peu de points. Comme l'estimation de densité par noyaux de Parzen, on utilise la même heuristique en prenant une étendue spatiale égale à deux fois la largeur de la fovéa (établie en pixels à l'écran). Nous utiliserons cette méthode combinée à une phase de correction manuelle pour l'initialisation de l'estimation des densités par mélange de gaussiennes.

5.2.1.3 Méthode d'estimation de densité par mélange de gaussiennes

L'algorithme de modélisation d'une distribution par Mélange de Gaussiennes ([Dempster77], [Biernacki06], [Dasgupta99], [Aitkin80]) permet l'estimation de la fonction de densité comme étant obtenue par la somme d'un nombre *a priori* connu de fonctions gaussiennes (les modes de la distribution). L'établissement de cette fonction de mélange s'effectue itérativement à partir de conditions initiales et jusqu'à convergence de la fonction de vraisemblance du modèle proposé. L'algorithme d'estimation des composants du mélange est l'algorithme « expectation-maximization ». La formulation usuelle de cette fonction de mélange est la suivante :

$$f(\vec{x}) = \sum_{k=1}^K p_k \cdot N(\vec{x}, \vec{\mu}_k, \Sigma_k)$$

avec :

- K le nombre *a priori* de gaussiennes,
- p_k la proportion du mode dans le mélange,

- $\vec{\mu}_k$ et Σ_k les paramètres (moyenne et matrice de variance covariance) du mode k.

Appliquée aux données oculométriques, cette formulation classique permet de capter les accumulations de points de fixations localisés spatialement. La présence de points de fixation plus diffus spatialement sera marquée par la formation d'un ou plusieurs modes gaussiens ayant une plus large dispersion. Il sera donc important de bien interpréter la matrice de variance-covariance de modes gaussiens pour savoir si l'on est en présence d'un mode représentant une zone d'intérêt spatialement bien délimitée ou d'un mode représentant des fixations distribuées spatialement plus largement. Pour bien marquer cette distinction, nous avons modifié l'algorithme « EM » pour intégrer dans le mélange un mode supplémentaire afin de rendre compte de l'existence, s'il y a lieu, d'une dispersion uniforme des données. Ainsi on utilisera une fonction de mélange « hybride » incluant un mode supplémentaire suivant une distribution uniforme :

$$f(\vec{x}) = \sum_{k=1}^K p_k \cdot N(\vec{x}, \vec{\mu}_k, \Sigma_k) + p_u \cdot \frac{1}{\Delta_u}$$

Avec :

- p_u la proportion du mode uniforme,
- Δ_u le support de la distribution uniforme (ici la surface de l'image).

Les proportions sont contraintes telles que

$$\sum_{k=1}^K p_k + p_u = 1$$

La difficulté de cette modélisation est qu'elle nécessite la connaissance *a priori* du nombre de modes de la distribution. Dans le cas de notre application, ce nombre est inconnu. La procédure qui est alors utilisée dans ce cas là est de mettre en compétition les différents modèles en faisant varier le nombre de modes et de sélectionner le modèle qui optimise un critère d'information comme par exemple le critère BIC (Bayesian Information Criterion) ou ICL (Integrated Classification likelihood) [Fraley98]. Dans notre cas nous avons appliqué une double sélection de modèles car nous avons mis en compétition les deux hypothèses de distribution: avec et sans mode uniforme.

Un élément fondamental de la modélisation par mélange de gaussiennes est le paramétrage de l'initialisation. Il est nécessaire de débiter les itérations de l'algorithme « EM » par une configuration initiale. Cette configuration comporte :

- Les valeurs des proportions : P_k
- Les positions des modes : μ_k
- La dispersion des modes : Σ_k

En pratique, usuellement les proportions sont choisies identiques, les matrices de variance-covariance sont choisies à l'identité ou proportionnelles à la matrice identité et les centres sont choisis aléatoirement. Puisque la modélisation finale va dépendre de ces paramètres de départ, une initialisation aléatoire fournira une modélisation différente pour des données identiques. On adoptera deux stratégies d'initialisation. La première est de choisir aléatoirement les positions et effectuer différents tirages, puis choisir le meilleur à la convergence sur la vraisemblance du modèle. La deuxième stratégie est d'utiliser un algorithme de détection de modes comme l'algorithme « mean

shift » pour fixer les positions initiales. Pour des raisons de temps de calcul on choisira la deuxième stratégie.

Cette méthode permet tout d'abord d'obtenir un modèle de distribution des points de fixations en prenant en compte la nature potentiellement aléatoire des données et donc de mettre en évidence les zones spatiales vers lesquelles l'attention visuelle s'est spécifiquement orientée. L'algorithme fournit de plus une description quantitative des modes donc permet une interprétation plus fine des distributions que la méthode des noyaux de Parzen. Les modes gaussiens de faible dispersion spatiale seront significatifs de la présence de zones spécifiquement vues et traitées. Au contraire, la présence de modes gaussiens à forte dispersion reflètera une nature plus aléatoire des données oculaires pouvant être interprétées comme une stratégie d'exploration de la scène plus globale. Enfin cette méthode est à la fois un moyen d'estimer la densité des points de fixation sous la forme d'une formulation analytique de la distribution en chaque point de l'espace mais elle est aussi un moyen de classifier les fixations, c'est-à-dire d'obtenir des groupes de fixations en fonction de la distribution spatiale globale. Les inconvénients de cette méthode résident dans la phase d'initialisation qui doit être bien réalisée et dans le temps de calcul pour la procédure de sélection des modèles en faisant varier le nombre possible de modes, avec et sans modes uniformes. C'est une méthode plus complexe qui ne se prête pas à la segmentation temporelle des données comme c'est le cas avec les noyaux de Parzen.

5.2.1.4 Conclusion sur les méthodes d'estimation de la distribution des fixations

Les deux méthodes d'estimation des fonctions de distribution spatiale des points de fixations apparaissent comme complémentaires. L'algorithme de modélisation par noyaux de Parzen permet de fournir rapidement une représentation visuelle descriptive de la densité des fixations sur l'image, ainsi que l'évolution temporelle de l'orientation de l'attention au cours de l'exposition. Cette méthode permet d'effectuer une estimation qualitative du nombre de zones fixées et de leurs positions, indiquant s'il est pertinent d'effectuer une analyse spatiale plus détaillée de la distribution des fixations. L'algorithme de modélisation par mélange de gaussiennes permet d'aller plus loin dans l'interprétation en fournissant le nombre de modes optimum et leur position, la présence ou non du mode uniforme, et l'orientation spatiale des modes. Enfin les résultats de la classification des fixations peuvent être exploités. Les modèles issus des mélanges de gaussiennes permettent ainsi l'observation spatiale globale de la distribution des fixations, et les noyaux de Parzen fournissent des indicateurs quant à la dynamique temporelle de l'attention. L'exploitation des résultats de la modélisation des points de fixation va consister à comparer ces modèles pour dégager le cas échéant les effets des conditions expérimentales associées.

5.2.2 Méthodes de comparaison des fonctions de distributions des fixations

5.2.2.1 Estimation de la similarité entre deux fonctions de distribution

La divergence de Kullback-Leibler [Kullback59] est une métrique de comparaison de deux distributions basée sur l'établissement de la vraisemblance de la fonction de densité $p(x)$ envers celle $q(x)$:

$$D(p|q) = \sum_{x \in X} p(x) \cdot \log\left(\frac{p(x)}{q(x)}\right)$$

On remarque cependant que la fonction D n'est pas commutative : $D(p|q) \neq D(q|p)$

On utilise alors la forme suivante (Bovick, Rajashekara [Rajashekara04]) :

$$D(pq)=1/(1/(D(p/q))+1/(D(q/p))$$

D'autres mesures de comparaison sont possibles. Sans vouloir être exhaustif, citons également la fonction de Kolmogorov-Smirnov (statistique 'KS') [Williams01] qui compare deux fonctions de répartition $P(x)$ et $Q(x)$:

$$KS= \max(|P(x)-Q(x) |)$$

Le test de même nom se déduit en utilisant la statistique de distance Kolmogorov-Smirnov.

La distance de Kullback-Leibler quantifie la distance entre deux modèles de densité tandis que la méthode de Kolmogorov-Smirnov, même si elle fournit un indice de comparaison entre modèles, est plutôt utilisée en tant que critère de décision de similarité des cartes de densité via leur fonction de répartition. Ces indices seront utilisés pour comparer les différentes conditions expérimentales.

5.2.2.2 Classification de la distance entre fonctions de densité

Lorsque l'on souhaite comparer la manière dont le regard a exploré un objet selon la consigne ou la modalité visuelle, il est intéressant de comparer les modèles de densité (modèles de densité de Parzen ou de Mélange de gaussiennes). La procédure suivante est alors suivie :

- Etablissement d'une matrice de distance de Kullback-Leibler calculée par comparaison deux à deux des modèles de densité
- Réalisation d'une représentation euclidienne des modèles par un algorithme « Multidimensionnal Scaling » afin d'observer l'organisation relative des modèles entre eux et si les axes de la représentation peuvent s'interpréter selon les facteurs expérimentaux
- Réalisation d'une Classification Ascendante Hiérarchique des modèles et confrontation des classes obtenues aux facteurs expérimentaux afin d'observer si la similarité entre modèles de densité est explicable par les conditions expérimentales

5.2.2.3 Utilisation de la fonction de vraisemblance des modes

La fonction de vraisemblance $L(x)$ est la probabilité d'apparition d'un échantillon

$X=(x_i, i=1:N)$ de fonction de densité $f(x)$ définie par un jeu de paramètres Θ :

$$L(X, \Theta)=\prod_{i=1}^N f(x_i, \Theta)$$

Plus le modèle sera adapté aux données, plus la vraisemblance sera importante. Dans des considérations numériques, c'est le logarithme de la vraisemblance qui est calculée. Pour notre application, si on tient compte de la durée de la fixation d_i comme pondération de la fixation x_i , la log-vraisemblance moyenne sera définie par :

$$L(X, \Theta)=\frac{1}{N} \cdot \prod_{i=1}^N p_i \cdot \log(f(x_i, \Theta))$$

avec p_i proportionnel à la durée de fixation.

On relève deux utilisations de cette fonction dans le contexte des données oculométriques. La première utilisation naturelle de la fonction de vraisemblance sera la confrontation des relevés oculométriques à leur modèles de densité afin d'observer si les modèles de densité de chaque condition expérimentale sont significatifs d'une tendance propre. La procédure est de calculer la vraisemblance moyenne de tous les points d'un essai sur chacun des modèles de densité de chaque condition expérimentale ; puis l'essai est attribué au modèle dont la vraisemblance est maximale. Connaissant le modèle auquel l'essai aurait du être attribué, il est ainsi possible d'établir le taux de confusion d'attribution selon le nombre de fois qu'un essai aura été mal attribué (attribué à un modèle d'une autre condition). Cette observation permet de dégager d'une part une indication de l'homogénéité interindividuelle intramodèles, mais aussi une indication de la force de guidage attentionnel des conditions expérimentales en fonction du type de mauvaises attributions.

Une seconde utilisation est de relever l'évolution temporelle de la vraisemblance en fonction du rang de la fixation. Par condition expérimentale, tous les essais sont regroupés et une modélisation de la densité est effectuée. Plus les points seront proches spatialement, plus la densité sera forte, plus la vraisemblance sera élevée ; et au contraire, plus les points seront éloignés, moins la densité sera élevée, et moins la vraisemblance sera forte. On peut ainsi observer si l'homogénéité du traitement spatial de l'image est corrélée au facteur temps (rang de fixation).

5.2.3 Méthode de classification et de regroupement des points de fixation

Afin d'analyser la répartition spatiale des données oculométriques, il est possible d'utiliser des algorithmes d'estimation de la densité, ce qui a été vu précédemment, mais aussi d'effectuer une classification des points de fixation pour identifier et isoler différentes zones ayant attiré le regard [Santella04]. Plusieurs méthodes peuvent être employées à cette fin, ayant chacune des spécificités et pouvant donc être utilisées dans des situations différentes.

L'algorithme « Meanshift » permet une première classification des points de fixation aux modes détectés auxquels ils contribuent le plus. On préférera néanmoins la classification issue de l'estimation par mélange de gaussiennes. En effet, cette méthode fournit une description plus précise des modes en tenant compte de leur dispersion spatiale par leur matrice de variance-covariance. Ainsi une phase de pré-classification est effectuée à partir de l'algorithme « Meanshift ». Cet algorithme fournit une liste de modes voisins et leurs points associés. Parce qu'on cherche à mettre en évidence les informations visuelles traitées, il peut être pertinent de fusionner des modes voisins, par exemple si le nombre de points de l'un d'eux est faible ou si les zones visuelles et sémantiques sur lesquelles reposent les modes ne se distinguent pas spécialement. Une phase de correction manuelle est ici effectuée pour confirmer la liste des centres d'initialisation de l'algorithme de Mélange de gaussiennes. A la convergence de l'algorithme « EM », on effectue une classification des points de fixation aux modes détectés, au sens du maximum de probabilité d'appartenance *a posteriori*.

La méthode de classification par mélange de gaussiennes est ainsi initialisée avec la liste de centres prédéfinis. Deux options sont alors possible : laisser la décision du nombre de clusters optimum à l'algorithme, celui-ci choisissant alors le modèle le moins couteux pour la classification, au risque de n'obtenir que deux clusters par exemple, ou bien fixer le nombre de clusters, l'algorithme ayant alors pour fonction tout d'abord de capter les fixations du mode uniforme, puis de repositionner les modes, et enfin de les orienter. C'est le type d'analyse exploitant la classification qui déterminera quels paramètres employer.

5.3 Analyse des zones d'intérêt

5.3.1.1 Définir les zones d'intérêt

L'analyse des zones d'intérêt consiste à corrélérer le comportement du regard et les informations visuelles (physiques et sémantiques) des zones qui ont attiré l'attention. Il est donc nécessaire pour cela de définir des zones spatiales qui seront comparées sur la base du comportement oculomoteur. Deux possibilités se présentent : définir les zones d'intérêt *a priori*, c'est-à-dire indépendamment de toute observation comportementale, sur une base sémantique, physique (texturale) ou géométrique, ou définir les zones d'intérêt *a posteriori*, basées sur les zones fixées sans prise en compte de la nature du stimulus visuel. Ces deux possibilités sont envisagées.

✓ *Décomposition a priori*

Au moins deux types de zones d'intérêt peuvent être définis *a priori* : les zones géométriques, les et les zones sémantiques.

- *Définition de zones géométriques*

Afin d'observer le comportement du regard en terme spatial, de manière systématique et dé-corrélé du stimulus, il est possible de décomposer la scène selon une grille géométrique, dont chaque case est identifiée individuellement. La dimension des cases peut varier de la taille de la fovéa jusqu'à la moitié de la largeur (ou hauteur) de l'image. Cette méthode a pour avantage d'être totalement indépendante de l'image et donc de pouvoir comparer des essais n'ayant pas été effectués sur le même stimulus.

- *Définition de zones d'intérêt sémantique*

Lorsque l'on dispose d'hypothèses préalables sur les zones d'intérêt sémantique de l'image et que l'on souhaite observer si ces zones reçoivent une attention visuelle spécifique, il est adapté d'établir une décomposition spatiale de l'image par un opérateur humain à partir de considérations sémantiques. Il peut s'agir par exemple d'identifier des zones porteuses de sens pour l'identification de la scène ou la résolution de la tâche, ou de définir l'emplacement de cibles. Chaque zone est alors encadrée par un polygone. Ces polygones peuvent être inclus ou concomitants. Dans le premier cas, les zones se superposent et chaque fixation peut être associée à plusieurs zones. Puisque l'information visuelle traitée ne se limite pas au point central de la fovéa, cette méthode permet de prendre en compte le fait que plusieurs zones puissent être traitées en même temps. Dans le deuxième cas, chaque partie de l'image n'est associée qu'à une seule zone, permettant de n'associer chaque fixation qu'à une seule zone, mais posant un problème lorsque la fixation est proche d'une frontière entre deux zones, la limite n'étant que géométrique et non pas visuelle.

✓ *Décomposition a posteriori*

La décomposition *a posteriori* est appliquée lorsque l'on ne dispose pas d'hypothèses préalables sur le déploiement de l'attention sur la scène ou lorsque l'on souhaite effectuer une analyse uniquement guidée par les données. Après avoir mesuré un ensemble de points de fixation, ceux-ci sont classifiés afin d'établir des groupes, qui seront considérés comme autant de zones d'intérêt.

- *Classification par un opérateur humain*

Cette méthode consiste à représenter l'ensemble des fixations superposée au stimulus. L'opérateur effectue des regroupements de fixations, en éliminant les points isolés, et en tentant d'associer les groupes avec des éléments picturaux significatifs. Cette méthode a pour principal défaut d'être soumise à la variabilité de jugement et d'estimation humaine, mais a comme avantage de permettre

un aperçu rapide des groupes et de leurs propriétés. Cette méthode ne peut en tout état de cause permettre de conclure sur le comportement oculomoteur mais peut aider à orienter la suite de la décomposition de l'image en zones d'intérêt.

▪ *Définition des zones d'intérêts par classification des points de fixation*

Cette approche algorithmique consiste à effectuer des regroupements de points de fixations en « clusters » (groupes) selon les dimensions spatiales, puis à analyser le comportement visuel associé à ces clusters. L'interprétation ne se fera qu'en toute fin d'analyse lorsque les clusters seront superposés à l'image.

La procédure que l'on suivra est la suivante :

- Phase 1 : estimation du nombre et de la position des clusters par classification
- Phase 2 : validation manuelle de la liste des centres. S'il existe une hypothèse *a priori*, il est possible de l'insérer ici sous la forme d'un centre complémentaire.
- Phase 3 : classification des points de fixation, c'est-à-dire association d'une fixation à un seul cluster, via la méthode de mélange de gaussiennes, commutable avec l'algorithme des k-mean [Forgy65]

Le principal intérêt est de pouvoir analyser de grandes quantités de données sans nécessiter d'hypothèses *a priori*.

5.3.1.2 Analyser le comportement du regard par zone d'intérêt

✓ *Etablissement de variables descriptives*

La première phase d'analyse des zones d'intérêt consiste à extraire des variables décrivant le comportement du regard. Les variables que l'on sélectionne sont les suivantes :

- Temps cumulé par zone : somme des durées des fixations associées à la zone
- Nombre d'entrées par zone : nombre de fixations associées à la zone
- Rang moyen de fixation (voir ci après score de Bradley-Terry) : indice indiquant si la zone est traitée plutôt tardivement ou plutôt rapidement dans l'ensemble des séquences. Plus le score obtenu par une zone est élevé, plus la probabilité que cette zone soit traitée tôt dans la séquence est élevée
- Durée moyenne des fixations par zone
- Amplitude de la saccade précédant et suivant la fixation de la zone
- densité moyenne portée par la zone : moyenne des densités des zones locales de toutes les fixations sur un modèle de densité. Plus cette variable est grande, plus les points sont concentrés, indiquant une forte attractivité locale de la zone.

A partir de ces variables quantifiées par zone, il va être possible de les comparer multi-dimensionnellement.

✓ *Comparaison des zones d'intérêts par les variables oculomotrices*

Cette étape a pour objectif de comparer les caractéristiques des zones d'intérêt en fonction des variables précédemment établies. Deux méthodes peuvent être appliquées : étude des corrélations des variables et du positionnement multidimensionnel des zones par une Analyse par Composante Principale, et classification/discrimination des zones par une analyse discriminante.

▪ *Méthode d'Analyse par Composante Principale*

Si l'on désire observer le comportement oculomoteur d'un faible nombre de zones (moins d'une dizaine) en fonction des variables précédemment décrites, la méthode la plus rapide est d'effectuer une simple Analyse par Composante Principale des zones selon ces dimensions. On obtiendra ainsi une description des positions relatives de chaque zone dans l'espace multidimensionnel. On cherchera ensuite à confronter la nature physique et sémantique des zones pour tenter d'expliquer cette distribution.

- *Méthode d'analyse par classification et discrimination*

Lorsque le nombre de zones est plus important et que l'on souhaite discriminer des groupes de zones selon le traitement visuel dont elles bénéficient, L'analyse factorielle discriminante ou une Classification Ascendante Hiérarchique est plus adaptée. On obtient ainsi des classes de zones pour observer si la nature sémantique des zones, rapportée à la consigne, permet d'expliquer le positionnement multidimensionnel.

5.3.1.3 Analyses complémentaires

- ✓ *Observation des transitions entre zones*

La matrice de transition a pour fonction de faire émerger les liens (perceptifs ou cognitifs) entre zones car elle met en évidence les occurrences de saccade d'une zone à une autre; elle s'établit en calculant le nombre de transitions entre chaque zone d'intérêt pour toutes les séquences de la condition expérimentale. En normant ces occurrences par le nombre total de saccades, on obtient la probabilité de transition du regard de la zone i vers la zone j . On observe s'il existe des stratégies d'exploration séquentielles récurrentes, explicables notamment par des effets de proximité spatiale, de similarité sémantique, ou de pertinence de tâche. Afin de vérifier si une transition est significativement surreprésentée, un test de Chi2 peut être appliqué si le nombre de d'occurrence minimum théorique par case est de 5 ou un test de Fisher pour des occurrences inférieures. Les transitions qui apparaissent comme surreprésentées sont analysées pour tenter de dégager le type de lien qui unit les deux zones fixées consécutivement. Après normage de la matrice, il est possible d'effectuer des tests de Mantel avec d'autres matrices de transition pour faire émerger les points communs et différences sous forme de matrices de dissimilarité. Notons qu'une représentation graphique par niveaux de couleurs indexés sur les occurrences permet de dégager rapidement visuellement les transitions sur- et sous- représentées.

- ✓ *Observation de l'ordre de fixation des zones*

- *Indicateur de rang*

Un indicateur de rang est calculé en attribuant pour chaque essai pour une zone donnée le numéro de rang de la première fixation qu'elle reçoit. En calculant la médiane de ces valeurs par zone pour l'ensemble des essais, on obtient un score, faible lorsqu'une zone est fixée tôt dans la séquence, et fort lorsqu'une zone est fixée tardivement. Même si cette approche est intéressante car très simple à mettre en œuvre, elle est limitée par la pertinence du calcul de la médiane, qui n'est pas représentative de la distribution des valeurs. C'est pourquoi nous nous intéressons au score de Bradley-Terry.

- *Test de Bradley-Terry : application aux zones de fixations*

Le score de Bradley-Terry [Bradley52] est une méthode issue de l'évaluation de produit par paires. Un juge attribue une note à chacun des deux produits de la paire. Lorsqu'il y a plusieurs produits, le juge effectue donc plusieurs évaluations et un produit reçoit donc plusieurs notes selon la paire dans laquelle il est jugé. De plus, plusieurs juges évaluent chaque paire. Une méthode d'analyse des notes doit ainsi être appliquée afin d'ordonner les produits selon leur occurrence de préférence ou de rejet. La méthode de Bradley-Terry permet ainsi de classer les produits selon la probabilité qu'un produit

soit préféré aux autres, en attribuant une valeur scalaire à chaque produit, plus cette valeur étant élevée, plus la probabilité que le produit soit préféré est forte.

On adapte ici à l'étude du rang des fixations. On considère que les zones d'intérêt sont les produits, et qu'une zone est « préférée » à une autre si elle est fixée plus tôt dans la séquence. On va donc établir pour chaque séquence une matrice de résultat synthétisant l'information suivante: lorsqu'une zone k est fixée par une fixation de rang n , alors la zone k est « gagnante » sur toutes les zones n'ayant jamais été fixées et perdante sur les zones fixées par les rangs $[1:n-1]$. Les matrices de toutes les séquences sont ensuite combinées. Le calcul du score de Bradley-Terry s'applique sur cette matrice totale, et fournit un score par zone, élevé lorsque la zone est souvent gagnante, et faible lorsqu'elle est perdante. Appliqué aux mouvements oculaires, une zone ayant un fort score de Bradley-Terry correspond à une zone fixée tôt dans la séquence et un faible score correspond à une zone fixée tardivement. L'intérêt de cette méthode est de prendre en compte la distribution des rangs de fixations par zone, et donc de mieux discriminer les zones selon le rang des fixations qui les traitent par rapport à un calcul de médiane des rangs des fixations associées à la zone.

5.3.1.4 Comparaison de séquences de fixations basée sur les zones d'intérêt

✓ *Distance de Levenstein*

L'objectif de ce traitement est d'observer si les mêmes zones d'intérêt sont fixées lors de différents essais. La distance de Levenstein [Levenshtein66] est une méthode de comparaison de séquence issue de travaux en génétique. Pour l'étude des séquences de fixations, elle nécessite de faire correspondre à chaque séquence de fixations une séquence de caractères de type « ababcd », où le n ème caractère correspond à un code attribué à la zone fixée par la n ème fixation. L'objectif de cette méthode est de quantifier la similarité entre deux séquences, afin de dégager les essais qui sont similaires en termes de « scanpath », c'est-à-dire d'ordre de fixation des zones. La comparaison des séquences se base sur le coût de transformation de la séquence A en séquence B, selon trois opérations de base : suppression, insertion, et substitution de caractères. Plus il y a d'opérations à effectuer, plus le coût est grand, plus la distance entre les deux séquences comparées est élevée. Deux résultats sont fournis, l'un prenant en compte l'ordre dans lequel les zones sont fixées, l'autre en négligeant l'ordre donc en ne se basant que sur la présence/ absence de zones. Cette méthode permet d'établir une matrice de similarité inter-essais, qui peut ensuite être employée pour tester les effets des conditions expérimentales sur les séquences de fixation.

Malgré sa rapidité d'exécution, cet algorithme possède plusieurs limites. Tout d'abord, selon le type de processus attentionnels impliqués, la pertinence de l'étude séquentielle est plus ou moins valide. En effet, dans le cas de la lecture par exemple l'attention est fortement sérielle ; on peut alors considérer que la zone fixée a une forte probabilité d'être traitée cognitivement et l'observation des séquences a donc un sens car il reflète l'intégration de l'information visuelle. Cependant, dans le cas de la perception des scènes réelles, le traitement de l'information visuelle ne se limite pas à la zone localement fixée, et la zone fixée n'est pas nécessairement traitée. L'ordre des zones fixées a donc peu de chance de représenter un processus d'intégration sériel des informations visuelles. La seconde difficulté de la méthode provient de son paramétrage. En effet, la pondération de chaque opération de manipulation de chaîne (suppression, substitution, insertion) influe directement sur la distance entre chaînes, mais il est difficile de relier ces coûts computationnels à des coûts cognitifs.

✓ *Data Time Warping*

La méthode « Data Time Warping » consiste à établir des similarités entre séquences temporelles, en faisant émerger des corrélations de répétition de motifs entre plusieurs signaux en fonction du temps [Keogh02]. L'algorithme fournit ainsi un indicateur de distance entre deux séquences temporelles ainsi qu'une séquence type qui reflète au mieux les points communs entre séquences en

limitant les variabilités locales. Le « data time warping » peut ainsi être appliqué aux séquences de fixation des zones d'intérêt, en faisant émerger un trajet temporel typique ainsi que les similarités de fixation des zones au cours du temps¹⁶.

✓ *Distance de Mannan entre deux séquences*

La distance de Mannan [Mannan95] est un algorithme de comparaison de distribution spatiale entre deux nuages de points. La somme de la distance de chaque point de la séquence P avec son plus proche voisin de la séquence Q et réciproque est établie. Cette méthode est donc insensible à l'orientation des nuages de points. Afin d'intégrer le fait que ces distributions puissent posséder une composante aléatoire, la distance sur les séquences mesurées est pondérée par la distance entre séquences dont les coordonnées des points ont été générés aléatoirement. Cette méthode peut être déclinée de deux manières : soit en comparant toutes les fixations effectuées dans un temps d'exposition similaire, soit en comparant toutes les fixations du rang 1 à n. Le choix dépendra du paradigme utilisé (durée d'exposition libre ou contraint par exemple).

5.4 Identification de stratégies d'exploration visuelle

Ce type de traitements consiste à caractériser la similarité des séquences de fixations par Analyse par Composante Principale. La première étape consiste à établir les variables descriptives du comportement du regard essai par essai : Nombre de fixations, Durée médiane des fixations, Amplitude moyenne des saccades, densités moyennes sur le modèle de densité, Etendue du nuage de points (rayons de l'ellipse englobant les points), etc. D'autres variables peuvent être ajoutés à cette liste non exhaustive selon les hypothèses que l'on souhaite vérifier. La seconde étape est d'effectuer une Analyse par Composante Principale des séquences mesurées selon les dimensions décrites ci-dessus. La première phase d'analyse consiste à observer les variables qui expliquent les axes ainsi que les corrélations. La seconde phase est l'observation de l'existence de groupes. On peut appliquer une Classification Ascendante Hiérarchique pour dégager des groupes, puis expliquer ces groupes par les dimensions de l'A.C.P. et ensuite par les conditions expérimentales (consigne, image,...).

5.5 Evaluation des facteurs prédictifs du comportement oculomoteur

Cette section recense plusieurs méthodes [Peters07] permettant de comparer des ensembles de valeurs afin de quantifier par exemple la qualité des prédictions des cartes de saillance sur le positionnement du regard.

La première méthode est la distance de Kullback-Leibler (mesure de dissimilarité) qui peut être utilisée pour mesurer la distance entre une carte de prédiction et la carte de densité des mesures. Pour chaque fixation, on établit un histogramme des valeurs de saillance aux points de fixations ainsi

¹⁶ Cette méthode a été testée et semble pertinente pour comparer des séquences de fixations. Cependant elle n'a pas donné de résultats exploitables sur nos données car celles-ci ne semblent pas comporter de motifs séquentiels spécifiques (attention visuelle non séquentielle)

qu'en des points positionnés aléatoirement. Puis la distance de Kullback-Leibler (ou entropie relative) est calculée entre ces deux distributions. Plus la distance de Kullback-Leibler sera faible, meilleure sera la prédiction. Cette méthode simple à mettre en œuvre a pour avantage d'être insensible à la reparamétrisation (par exemple l'application d'une fonction monotone continue non linéaire sur la carte de prédiction), mais a pour inconvénient de ne pas avoir de limites supérieures bien définies, donc de diverger vers l'infini.

La seconde méthode est celle des percentiles (mesure de similarité). Il s'agit aussi d'une distance qui est insensible à la reparamétrisation, mais elle ne diverge pas vers l'infini. Une grande valeur du percentile signifie que la prédiction est bonne. Le calcul se fait en établissant pour chaque point de fixation le nombre de lieux de la carte de prédiction qui ont une valeur inférieure à celle du lieu de fixation. La moyenne et la déviation standard sont ensuite calculées.

La troisième méthode est nommée « Normalized Scanpath Saliency » (NSS) [Peters05]. Elle permet de comparer les lieux de fixations mesurées avec les prédictions des cartes de saillance. Chaque carte de saillance est centrée et normée. Pour chaque lieu de fixation, on mesure la valeur de déviation standard de la saillance, et on compte le nombre de valeurs de la carte qui lui sont inférieures. Plus la prédiction est bonne, plus il y aura de valeurs inférieures, plus la valeur NSS sera grande. Si au contraire la prédiction est totalement décorrélée de la mesure, alors la valeur NSS sera faible. Contrairement à la distance de Kullback et aux percentiles, la méthode NSS est sensible à la reparamétrisation.

Une quatrième méthode permet de vérifier si une carte de prédiction prédit bien le positionnement des fixations. On va chercher à vérifier si les valeurs de prédiction mesurées aux lieux des fixations sont significativement différentes de valeurs mesurées à partir d'autres distributions. La première distribution référence est un ensemble de points positionnés aléatoirement dans l'espace, et la seconde distribution est un ensemble de points de fixations mesurées dans des conditions expérimentales différentes. Un test de Kolmogorov-Smirnov ou une analyse de variance va ensuite permettre de définir si les valeurs de prédictions aux lieux de fixation sont significativement différentes des valeurs des deux autres distributions. Si cela se vérifie, alors la carte de prédiction démontre effectivement que le regard se pose sur des zones spécifiques.

Pour l'estimation du caractère prédictif de plusieurs facteurs, nous proposons d'employer une méthode d'estimation de contributions d'hypothèses par mélange additif basé sur l'algorithme d'expectation-maximization précédemment abordé. Chaque hypothèse de guidage du regard est associée à une carte de probabilité spatiale. Un mode gaussien uniforme permet de capter les fixations distribuées aléatoirement. L'algorithme fournit une valeur de contribution de chaque carte de probabilité (chaque hypothèse) à la distribution mesurée des fixations. Cette méthode sera développée dans l'avant dernier chapitre.

5.6 Synthèse

Plusieurs niveaux de traitements ont été présentés, et correspondent à une démarche d'approche à l'analyse du comportement oculomoteur. Tout d'abord les données de base ont été transformées, partant d'un signal brut pour obtenir des séquences de points de fixation. Puis des variables descriptives des séquences de fixations ont été établies afin de décrire le comportement oculomoteur en terme spatial et temporel et des cartes de densité spatiale et spatio-temporelles sont estimées. Les modèles et les variables sont ensuite traités par des méthodes de classification et d'analyse multidimensionnelle pour dégager des tendances dans les processus attentionnels. Trois phases

successives de traitement sont ainsi proposées, de l'étude globale à l'étude locale : Exploration globale par les cartes de densités, établissement de zones d'intérêt et analyse du traitement visuel de ces zones

Ce chapitre a présenté les variables et méthodes utilisées dans le cadre de ce travail, sans qu'il ne s'agisse d'une revue exhaustive de toutes celles qui peuvent être employées dans l'analyse de ce type de données. De même, ne sont pas exposés ici les traitements spécifiques appliqués dans certains de nos paradigmes (ex : comparaison par paire) afin de clarifier l'approche méthodologique (et non le test d'hypothèses spécifiques). Notre travail s'inscrivant dans le domaine de la perception visuelle des scènes réalistes et de l'évaluation d'objets, les traitements et les procédures sont orientés afin de favoriser une approche statistique, limitant les effets de la variabilité idiosyncrasique de chaque essai. Enfin les différentes représentations graphiques des données ne sont pas détaillées ici car elles sont dépendantes des paradigmes employés.

A partir de ces différentes méthodes va pouvoir être effectuée l'analyse des comportements oculomoteurs mesurés lors de l'évaluation d'objets manufacturés photographiés. Les éléments théoriques qui ont été présentés précédemment nous montrent que les processus attentionnels sont nombreux, dynamiques, et peu répétables. De plus, l'évaluation est l'aboutissement de mécanismes de traitement de l'information à différents niveaux cognitifs et sous l'influence de différents biais. L'élaboration de nos protocoles expérimentaux doit ainsi prendre en compte cette forte variance comportementale et la nature incertaine (voire instable) des informations sur lesquelles repose l'évaluation. Le chapitre suivant présente une expérimentation d'évaluation monadique des habitacles de véhicules.

Chapitre 6

Expérimentation 1

« paradigme d'évaluation monadique »

6.1 Matériel et méthodes

6.1.1 Objectifs

Evaluer visuellement un objet consiste à recueillir des informations sur ce dernier afin de juger ses différentes propriétés et d'exprimer un niveau global de satisfaction. L'évaluation monadique est une situation dans laquelle le participant ne compare pas l'élément étudié à d'autres éléments de la même catégorie, mais à des références encodées en mémoire, en fonction de ses préférences et de ses attentes. Malgré le fait que le jugement d'un objet en situation monadique ait tendance à produire un jugement sur-favorable à celui-ci ([Robinson85], [Sanbonmatsu98]), cette situation est intéressante car on la rencontre tous les jours. Cette première expérimentation a pour objectif de vérifier si les participants évaluent le produit tel que l'a défini le concepteur. Dans l'affirmatif, y a-t-il un comportement oculomoteur typique de chaque condition expérimentale ? Si non, y a-t-il une homogénéité entre les participants, tant en terme d'exploration visuelle que d'évaluation. On s'intéresse enfin à identifier les principaux facteurs endogènes (provenant de propriétés cognitives) et exogènes (provenant de propriétés perceptives) qui semblent guider le regard. Cette étude est menée en élaborant un paradigme d'évaluation monadique du produit automobile sur photographies à échelle 0.8 (échelle contrainte par la taille d'écran, l'objectif étant de se rapprocher de l'échelle 1) avec des participants non experts du monde automobile.

6.1.2 Matériel

Tel que cela a été présenté précédemment, les consignes sont choisies dans le but de produire des jugements différents entre les modalités visuelles de chaque produit. Les produits choisis (voir chapitre « cadre général ») sont la Peugeot 207 sous la déclinaison « sport » et « standard » et la Citroën C6, sous la version « Blanche » et « Noir ». Pour le produit 207, les modalités visuelles sont définies comme « sport » et « standard » par les concepteurs. La première consigne est d'évaluer le niveau de sportivité. Si les intentions concepteurs sont bien reconnues, les deux modalités ne doivent pas obtenir les mêmes notes. La deuxième consigne est d'évaluer le niveau de gamme. Pour le produit C6, dont les modalités visuelles sont soit de couleur dominante noir soit blanche. La première consigne est d'évaluer le niveau de raffinement du produit et la seconde consigne est d'évaluer le niveau de gamme. Ces deux consignes se basent sur des critères plutôt ambiants (informations visuelles globales), de la même manière que les deux modèles de C6 diffèrent par des attributs visuels globaux.

Des photographies couleur à très haute résolution (3200*2400) et à haute dynamique de luminance sont effectuées sur les produits prédéfinis, en adoptant un point de vue et une focale strictement identique pour chaque modalité ; les images sont ensuite recadrées à l'échelle 0,8 (Planche 1).



Planche 1 Stimuli employés dans l'expérimentation monadique

Une pré-expérimentation qualitative teste les consignes sur des participants non entraînés avec différentes durées d'exposition. Il apparaît que 8 secondes est la durée moyenne à laquelle la tâche est correctement effectuée.

Avant exposition, il leur est précisé qu'ils vont juger des habitacles de véhicules, que la durée d'exposition est de huit secondes, et qu'ils devront ensuite fournir leur réponse. On leur demande : « observer la scène attentivement et évaluez le caractère sportif, (ou haut de gamme), (ou raffiné) de l'habitacle ». La séquence est ensuite composée d'une diapositive contenant un cercle de 5° d'angle visuel de large s'affichant au centre de l'image sur fond noir pendant une seconde, puis une diapositive contenant une image de bruit blanc pendant une demi-seconde. La diapositive contenant la photographie du produit est alors exposée pendant 8 secondes. Les participants donnent ensuite leur réponse de jugement sur une échelle en 9 points de pas du tout à très « sportif » (ou « haut de gamme » ou « raffiné ». Enfin, ils verbalisent à haute voix en désignant sur un tirage papier les zones qui ont aidé à leur décision.

8 groupes de 10 participants sont constitués, chaque groupe étant associé à une condition expérimentale (2 consignes par produit * 4 produits à évaluer).

6.2 Résultats

6.2.1 Introduction : Jugement des produits

La première phase de traitement des données expérimentales consiste à observer les notes obtenues sur les différents objets afin de dégager s'il existe une différence de jugement ou non entre eux. Ces résultats seront ensuite confrontés aux mesures oculométriques. Chaque version d'habitacle était jugée en déplaçant sur une échelle en 9 pas un curseur, indiquant une note de 1 à 100.

6.2.1.1 Produit 207

La spécificité du produit 207 est que la différence entre les deux modalités visuelles (version « sport » et « haut de gamme ») porte sur des attributs spécifiques composant les habitacles. Le Tableau 2 présente les notes sur 100 obtenues par chaque version du véhicule 207.

	207 sport	207 standard
Niveau de sportivité	60	21
Niveau de gamme	56,5	42,8

Tableau 2 Moyenne des notes sur 100 des habitacles par consigne et par modalité visuelle de 207

Un test de Student sur groupes non appariés montre que les notes qu'obtiennent les deux modalités visuelles (les deux versions d'habitacle) sont significativement différentes pour la consigne de sportivité ($\alpha=0.03$) mais pas pour la consigne de niveau de gamme ($\alpha>0.05$). Ce test montre de plus que les notes obtenues par la modalité 207-sport ne sont pas significativement différentes entre les deux consignes, tandis que la modalité « standard » est significativement mieux notée pour la consigne « niveau de gamme » que pour la consigne « sportivité » ($\alpha=0.04$).

Ces résultats montrent que les notes obtenues par les deux modalités visuelles sont en accord avec les prédictions des concepteurs, c'est-à-dire que la 207-sport est considérée comme plus sportive que la 207-standard. Les deux habitacles ne sont cependant pas aussi fortement discriminés sur le critère niveau de gamme, mais il y pourrait y avoir une inférence des attributs sportifs envers le niveau de gamme car c'est l'habitacle considéré comme le plus sportif qui est aussi le plus « haut de gamme ».

Puisque l'on observe des différences de jugement entre les deux habitacles, on va tenter de dégager s'il existe aussi des différences dans la prise d'information et dans les processus attentionnels.

6.2.1.2 Produit C6

La spécificité du produit C6 suivant ces deux finitions est de ne se distinguer que sur des propriétés « ambiantes » de nature « signal », et que les consignes ne sont pas orientées pour la vérification d'une intention concepteur particulière. Le Tableau 3 présente les moyennes de notes obtenues par chaque habitacle pour chaque consigne.

	C6 noire	C6 blanche
Niveau de raffinement	68	81,3
Niveau de gamme	75,5	76,6

Tableau 3 Moyenne des notes sur 100 des habitacles par consigne et par modalité visuelle de C6

Un test de Student non appareillé montre que les deux modalités visuelles sont significativement différemment notés pour la consigne « niveau de raffinement » ($\alpha < 0.05$) mais pas pour la consigne « niveau de gamme ». Ce même test montre que la C6-Blanche n'obtient pas les mêmes notes entre les deux consignes, contrairement à la C6-noire qui obtient des notes assez similaires. Ainsi, La C6-Blanche est mieux notée que la C6-noire en termes de raffinement, et la C6-Blanche est considérée plus raffinée que haut de gamme.

Les notes moyennes d'évaluation du niveau de gamme des deux habitacles de C6 sont identiques. Deux hypothèses peuvent être émises. Soit aucune des deux modalités visuelles ne contienne d'attributs qui permettent de répondre à la consigne, soit que ces attributs sont présents dans les deux modalités.

La confrontation de ces résultats aux mesures oculométriques va nous permettre d'observer si la différence de jugement entre les deux modalités visuelles pour la consigne niveau de raffinement et la non discrimination pour la consigne niveau de gamme se reflète dans la prise d'information visuelle.

6.2.2 Analyse des données oculométriques : fonctions de densité spatiale des fixations

6.2.2.1 Introduction

La première phase du traitement des données oculométriques a pour objectif de fournir une description globale spatiotemporelle du déploiement de l'attention sur la scène visuelle. La pertinence d'étudier en détail les relations entre comportement du regard et les différentes zones de l'image dépendra des résultats de ces premières observations.

Cette étape se décompose en deux phases :

- Analyse qualitative descriptive des données oculométriques pour dégager les tendances de comportements oculomoteurs de chaque condition expérimentale
- Comparaison des indicateurs descriptifs entre les différentes conditions expérimentales pour quantifier les effets des facteurs expérimentaux (consigne, modalité visuelle).

6.2.2.2 Analyse descriptive de la densité de Parzen

On pose l'hypothèse pour cette analyse que les participants ont un comportement oculomoteur homogène pour une même consigne et une même image. On va donc chercher à observer s'il existe des spécificités des distributions spatiales des fixations pour chaque condition expérimentale. Pour cela, on va établir la fonction de distribution des densités de fixations de Parzen. Ce traitement est le premier à être établi car

- il est déterministe (le résultat de la fonction est le même pour des données d'entrée identiques)
- il permet d'observer une tendance temporelle de l'évolution de l'attention visuelle sur l'image
- il s'agit de résultats visuels synthétiques

Le calcul de la fonction de distribution spatiale des fixations par la méthode des densités de Parzen est effectué pour chaque condition expérimentale, c'est-à-dire en regroupant tous les essais effectués sur la même modalité visuelle avec une consigne identique.

✓ *Résultat pour le produit 207*

La figure suivante présente les densités de fixation pour les différentes conditions expérimentales.

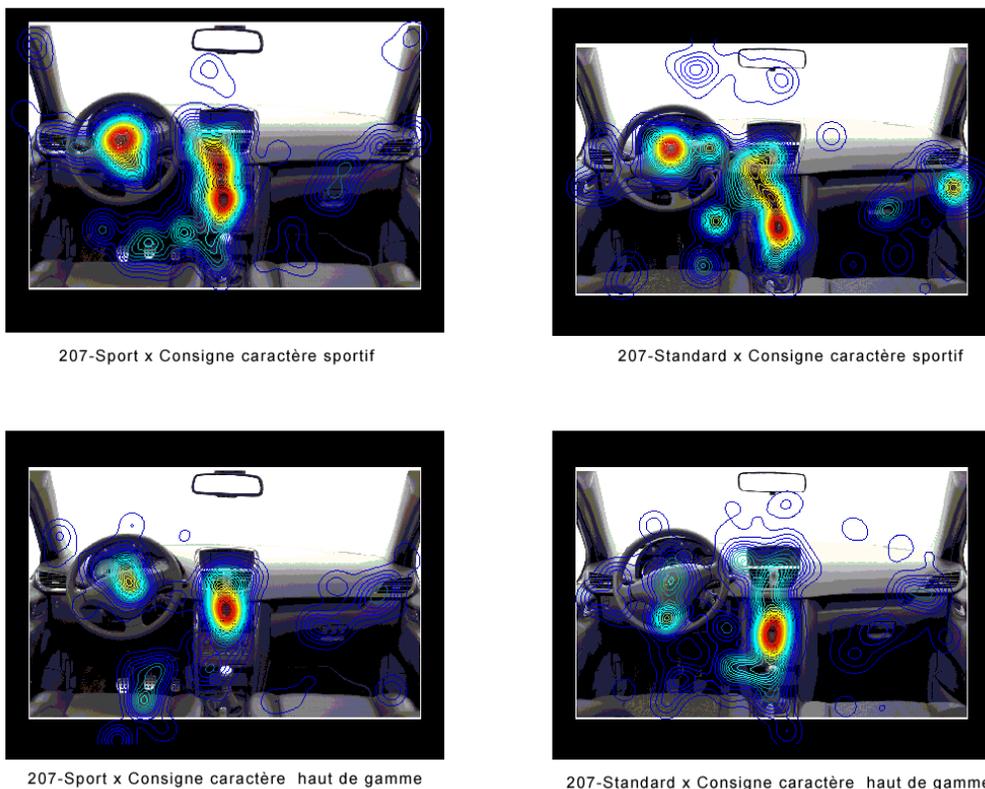


Figure 14 Densité de Parzen des fixations pour chacune des 4 conditions expérimentales du produit 207

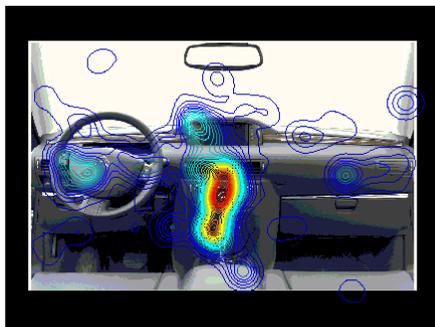
Sur la Figure 14, les densités des fixations de deux modalités visuelles (207-sport & 207-standard) d'un même produit (207) observées sous deux consignes (C1 – caractère sportif- & C2 – caractère haut de gamme) sont représentées.

On peut observer pour le cas « 207-sport consigne caractère sportif » que les courbes de niveaux de densités sont très resserrées au niveau du centre du volant et de la console centrale. Elles entourent le pédalier et le levier de vitesse même si elles y sont peu denses. Cela indique que le regard s'est positionné de nombreuses fois sur le volant et la console centrale, et peu de fois mais de manière précise sur le pédalier et le levier de vitesse.

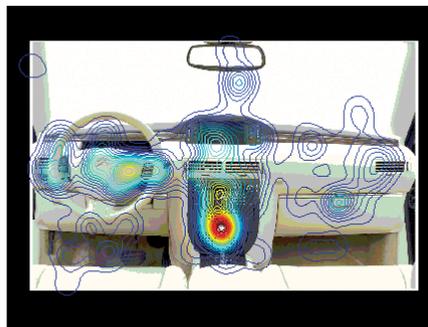
Qualitativement, les quatre modèles de densités semblent très similaires par la rapide apparition d'un mode principal sur le centre de la console centrale, qui s'avère aussi être le centre de l'image. Cependant on note pour la consigne sportivité la présence d'un second mode principal sur la zone volant/logo/compteur. On observe pour le produit 207-sport que la zone pédalier est spécifiquement traitée, sachant qu'il existe dans cette zone un pédalier spécifiquement visible, contrairement au produit 207-standard où le pédalier est presque invisible.

✓ Résultats pour le produit C6

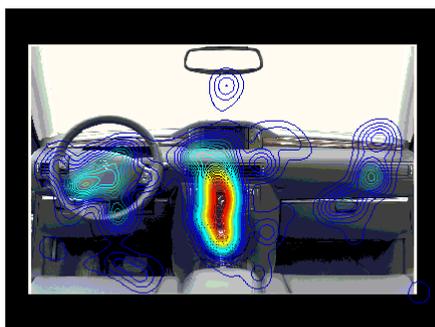
La figure suivante présente les densités de fixations pour le produit C6.



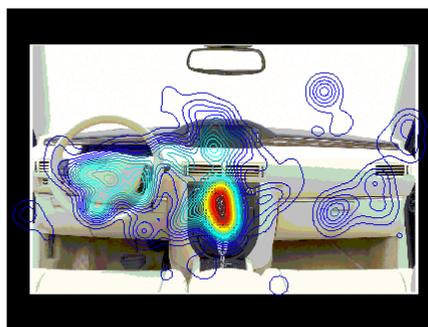
C6-Noire x Consigne caractère raffiné



C6-Blanche x Consigne caractère raffiné



C6-Noire x Consigne caractère haut de gamme



C6-Blanche x Consigne caractère haut de gamme

Figure 15 Densité de Parzen des fixations cumulées pour chacune des 4 conditions expérimentales du produit C6

Sur la Figure 15, les densités des fixations mesurées sur deux modalités visuelles (C6-Blanche & C6-Noire) d'un même produit observées sous deux consignes (C3 – caractère raffiné- & C4 – caractère haut de gamme) sont représentées. Les points suivants se dégagent. Tout d'abord on observe une forte similarité de la distribution de l'attention entre les 4 cas ; il existe ainsi un important mode de densité au centre de la console centrale qui est aussi le centre de l'image, et l'observation de la zone volant/logo/compteur est intense. On note que l'attention semble se porter sur une plus grande surface de l'image que pour la 207, comme s'il existait peu d'attracteurs attentionnels locaux, produisant une prise d'information « ambiante » plutôt que « focale ». La consigne ne semble pas orienter spécifiquement le regard.

On note pour les 8 conditions expérimentales que les zones centrales de la scène visuelles sont fortement observées ; plusieurs hypothèses peuvent l'expliquer :

- biais d'observation centrale de l'image
- zone console centrale fortement porteuse de sens pour ce type de produit
- point de vue d'observation optimisé et privilégié pour traiter la diversité des informations en présence dans l'image.

Il est donc intéressant d'observer si la dynamique attentionnelle s'organise autour de ce point central.

6.2.2.3 Evolution de la densité au cours du temps

Les résultats détaillés de chaque condition expérimentale sont présentés en annexe (section 1.1.1). La Figure 16 représente l'évolution de la densité de Parzen par fenêtrage temporel pour la condition « 207-sport consigne sportivité ».

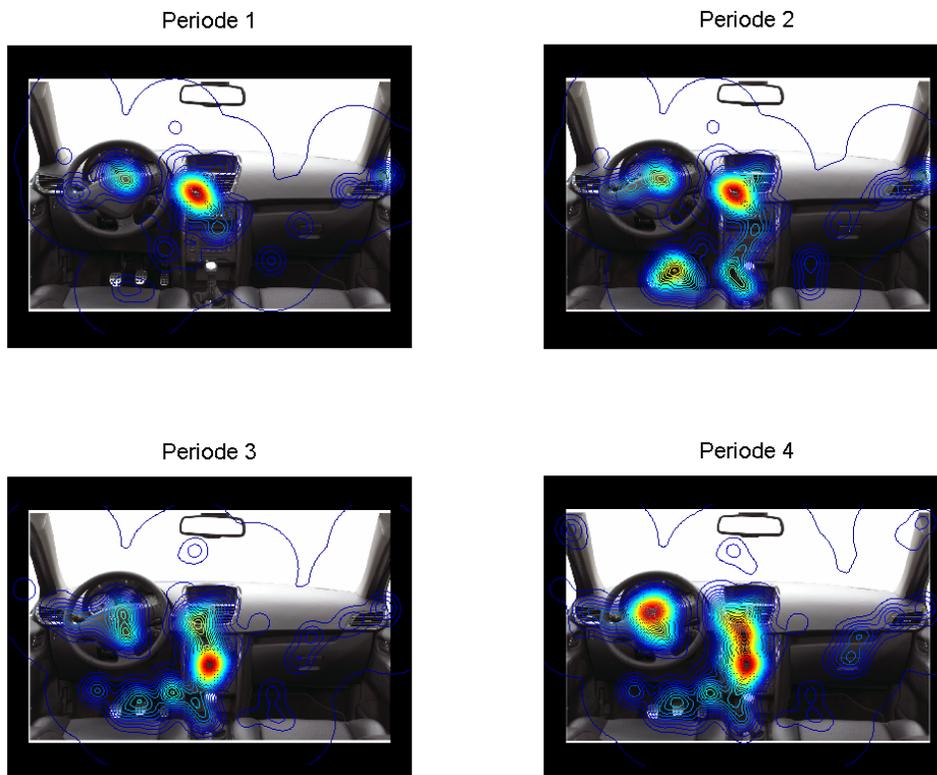


Figure 16 Evolution temporelle de la densité de Parzen par pas de 2 secondes pour le cas 207-sport/consigne sportivité

Les densités de fixations de la Figure 6-3 ont été calculées par cumul des fixations au cours du temps lors de l'évaluation de la sportivité sur la 207-sport. La période 1 correspond aux fixations mesurées dans les deux premières secondes, la période 2 correspond aux fixations des 4 premières secondes, la période 3 aux 6 premières secondes et la période 4 aux 8 premières secondes. Contrairement au calcul par fraction (période 1 : 0-2s, période 2 : 2-4s, etc.), cette méthode permet de dégager l'évolution de la surface fixée par le regard, et le renforcement ou la disparition de modes existants, aboutissant à la situation finale. Ainsi, sur la figure 3-3 on observe que le regard reste centré pendant les deux premières secondes dans l'image au niveau du point de fixation initial (mode important au centre de l'image – courbes de niveaux de couleurs rouge). Il existe cependant un mode seconde (moins intense : peu de fixations) sur la zone logo/compteur. Il n'y a donc pas, hormis cette zone logo/compteur, de zones attirant fortement le regard dans les premiers instants. Puis apparaissent dans la période suivante des modes secondaires sur le levier de vitesse et sur la console centrale. Dans la période 3 le mode central se renforce et se déplace, le regard semblant explorer plus en détail la console centrale. Enfin dans la dernière période le regard s'oriente vers les compteurs.

Pour toutes les conditions expérimentales, il existe un mode central présent dès les deux premières secondes, qui se renforce au cours du temps. On remarque que lorsqu'il existe des attributs spécifiques tels que des pédales ou levier de vitesse brillants, ceux-ci attirent le regard, à partir de la deuxième période. Enfin, sur C6, on observe que quelle que soit la période, le regard explore une

grande surface de l'image. Chaque cas possède ses propres spécificités d'orientation temporelle du regard, sans qu'il soit évident, par cette analyse descriptive, de dégager des effets propres aux conditions expérimentales où à des groupes de sujets. Cependant, les observations de la dynamique du regard semblent confirmer le fait que le regard favorise une position centrale pour traiter l'image, et qu'à l'issue des deux premières secondes, de nouveaux modes apparaissent malgré le fait que l'exploration est fortement centrale. Ce délai serait donc celui qui, en l'absence d'attracteurs attentionnels immédiats, permettrait un traitement global suffisamment abouti pour déplacer le regard sur des zones locales.

L'intérêt de cette analyse est d'obtenir des indications sur le caractère dynamique ou statique du regard au cours de la séquence. On constate dans cette situation qu'il existe différentes zones attirant le regard, mais que les fonctions de distribution de densité sont assez similaires quelles que soient les conditions expérimentales. L'interprétation de l'analyse plus détaillées du traitement des différentes zones de la scène visuelle devra donc être effectuée avec une grande prudence.

La densité de Parzen est une représentation descriptive de la distribution des fixations. Cette méthode ne fournit pas de critères permettant de comparer les essais. Ainsi, afin d'obtenir des informations complémentaires sur la spécificité des différentes distributions de fixations, on va appliquer sur les données une modélisation par la méthode des mélange de gaussiennes.

6.2.2.4 Modélisation de la distribution des fixations par Mélange de Gaussiennes

Comme pour l'estimation de la densité de Parzen, on cherche ici à modéliser l'ensemble des fixations mesurées pour tous les essais d'une même condition expérimentale (image x consigne), en considérant que ces conditions expérimentales sont suffisamment influentes sur le comportement du regard pour permettre une homogénéité inter-essais.

Cette méthode de modélisation a pour avantage de fournir des critères de comparaison des modèles de distribution des fixations, de dégager des modes distincts de fixations, où chaque fixation n'est associée qu'à un mode, et donc de fournir une estimation analytique de la densité de chaque mode et de la densité totale. Enfin, cette méthode intègre dans le modèle une composante aléatoire, permettant d'éliminer les fixations distribuées uniformément et donc de mieux dégager les cas où l'exploration est globale (ambiante) et où le regard s'attarde sur des zones spécifiques (focale).

Tandis que la section précédente a mis en évidence l'existence de modes de fixations (pics de densité), cette étape va consister à définir dans le modèle de densité le nombre, la position et l'intensité de chaque mode. Plusieurs types d'initialisations sont possibles. Dans le cas présent, nous choisissons d'établir un ensemble de barycentres par méthode des Meanshift, permettant d'initialiser le modèle des mélanges de gaussiennes autrement que par une fonction aléatoire. Plusieurs calculs de barycentres sont effectués jusqu'à obtenir un groupe de barycentres stable ; la stabilité de ce groupe est prédit par le fait que la densité de Parzen montre l'existence de modes localisés et non un traitement uniforme de l'image. On permet ensuite à un opérateur humain de supprimer ou fusionner des modes qui seraient associés à une même zone visuelle ou sémantique. Cette opération décisionnelle est appliquée malgré son caractère subjectif car on considère que les distributions des fixations sont liées aux informations visuelles sémantiques, que perçoit l'humain et non la machine, et enfin car il s'agit d'une procédure d'initialisation, et non du modèle définitif, qui sera, lui, fourni par l'algorithme de mélange de gaussienne. On obtient de cette manière les barycentres qui vont servir à initialiser la modélisation par mélanges de gaussiennes.

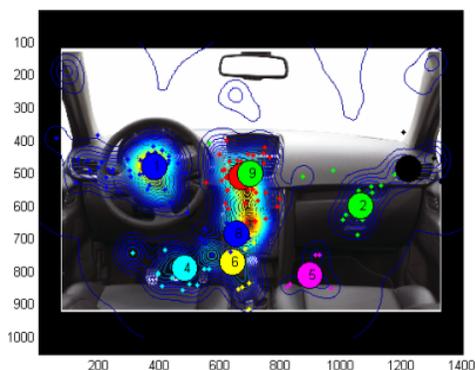


Figure 17 Résultat de la classification des points de fixation par la méthode « Meanshift » superposée à la fonction de densité de Parzen

La Figure 17 représente les barycentres de fixations fournis en sortie de Meanshift, superposés aux courbes de densité de Parzen, pour la condition 207-sport / consigne sportivité. Le mode rouge '3' sera éliminé car superposé au mode vert '9' : il n'est pas pertinent de distinguer deux zones non discriminées par la vision fovéale. Le mode bleu '8' sera fusionné avec le mode jaune '6' car très proches, sachant que si les deux modes étaient en fait bien distincts, l'algorithme des mélanges de gaussiennes les dissociera. Enfin le mode magenta '5' est éliminé car associé à un très petit nombre de points ; là aussi l'algorithme de modélisation recréera le mode si celui-ci s'avère consistant.

A l'issue de cette initialisation nous obtenons une liste de centres de mode définis par leurs coordonnées. Ces centres vont ensuite être ajustés par l'algorithme « EM » au cours des itérations d'apprentissage.

La modélisation par mélange de gaussiennes de la distribution spatiale des fixations est alors effectuée. En faisant varier le nombre *a priori* de modes, avec ou sans mode uniforme, les modèles sont mis en compétition. Le meilleur modèle est sélectionné au sens du critère d'information ICL. Pour le modèle sélectionné, afin de caractériser et comparer les effets des conditions expérimentales, on retiendra :

- L'existence ou absence d'un mode uniforme accaparant les fixations liées à une composante non déterministe de la distribution
- les paramètres du modèle : nombre, positions, orientations et intensités des modes

Les résultats de la modélisation sont présentés en annexe en section 1.1.2.

✓ Modélisation pour le cas 207-sport/consigne sportivité

On illustre ici la sélection du modèle dans le cas de la 207-sport sous la consigne « caractère sportif ». Pour la condition expérimentale « 207-sport, consigne caractère sportif », on observe les caractéristiques suivantes au modèle.

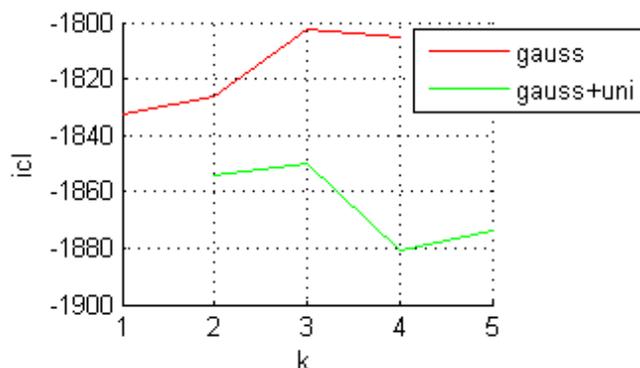


Figure 18 Evolution du critère d'information avec et sans mode uniforme, pour les modèles de 1 à 5 modes.

La Figure 18 présente l'évolution du critère d'information « ICL » en fonction du nombre de modes gaussiens contenus par le modèle et en fonction de la présence ou absence d'un mode uniforme. Plus le critère ICL est élevé, plus le modèle est de qualité (mieux la variabilité de la distribution est intégrée dans les différents modes ??). On observe ici que le meilleur modèle est constitué de 3 modes gaussiens sans mode uniforme. C'est donc cette configuration qui représente au mieux la distribution des fixations.

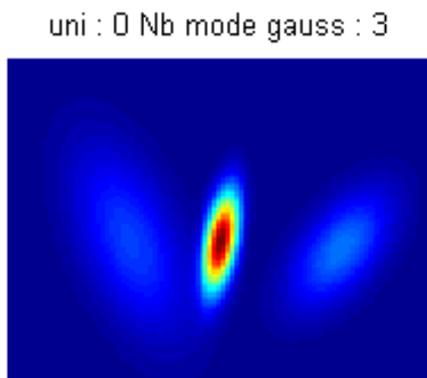


Figure 19 Densité spatiale des fixations définie par les modes gaussiens

La Figure 19 représente la densité spatiale des 3 modes gaussiens (et la fonction de distribution de probabilité des fixations) définis au préalable. On en tire que le mode central est le plus intense, c'est-à-dire que le regard s'est souvent positionné sur cette zone spatiale. De plus, ce mode central est moins étendu que les deux autres, signifiant que le regard y était peu dispersé. Cette zone correspond au lieu original du point de fixation pré-exposition, mais aussi au centre spatial de l'image

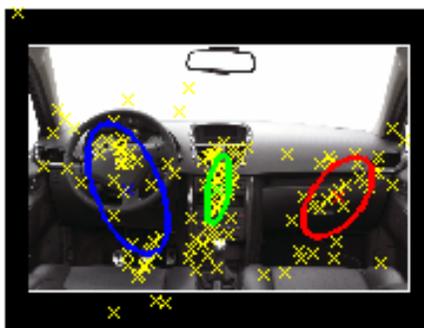


Figure 20 Groupement des fixations par modes gaussiens

La Figure 20 représente les contours à un écart-type des modes gaussiens superposés aux fixations, indiquant la position et la forme de chacun des modes, mais aussi la manière dont les fixations sont associées aux modes. On remarque ici pour ce cas que les fixations sont effectivement très concentrées pour le mode central, qu'elles semblent constituer deux sous-groupes pour le mode de gauche, et qu'elles sont plutôt éparpillées pour le mode de droite.

✓ *Synthèse - Existence d'un mode uniforme*

Aucune des modélisations n'inclut de mode uniforme. Cela signifie qu'il n'y a pas de fixations qui pourraient correspondre à une distribution aléatoire uniforme des points sur l'image. Un mode gaussien uniforme qui couvrirait la totalité de la surface de l'image ne capterait donc aucune fixations. Cela indique que les processus guidant le regard ne sont pas de nature aléatoire, mais aussi que l'exploration spatiale de la scène est structurée, et non distribuée équitablement sur toutes les zones de l'image.

✓ *Synthèse - Nombre de modes*

Pour chaque condition expérimentale, on relève le nombre de modes qu'inclut le modèle de mélange de gaussiennes. Le Tableau 4 synthétise ces informations.

nombre de modes	207-sport	207-standard	C6-noire	C6-blanche
Consigne 1	4	6	1	4
Consigne 2	3	1	4	1

Tableau 4 Nombre de modes pour chaque condition expérimentale consigne-modalité visuelle.
Consigne 1 (sportivité ou raffinement selon le produit 207 ou C6 respectivement) Consigne 2 (niveau de gamme)

Le nombre de modes est très variable selon les conditions expérimentales. La présence du concept de sportivité, soit dans l'image, soit dans la consigne semble induire une exploration en zones localisées. Ainsi, le produit 207 sous la consigne sportivité est modélisé par plusieurs modes. Pour le produit C6, l'inversion du nombre de mode selon la consigne entre les deux typages ne permet pas de conclure sur l'effet du facteur consigne ou du facteur image sur le déploiement de l'attention.

Lorsque la distribution est modélisée par un unique mode, il peut s'agir des conséquences :

- d'un fort biais d'observation centrale
- d'un faible besoin d'informations locales pour la résolution de la consigne
- de l'absence d'informations pour la résolution de la consigne
- d'une maximisation de l'acquisition de l'information depuis ce point de vue

Lorsque le modèle comporte 3 modes, l'un est positionné sur la zone des instruments de conduite, un autre est le mode sur la console centrale, le troisième est sur le côté passager.

Lorsque le modèle comporte 4 modes, la répartition est similaire à la répartition en 3 modes, à la différence que le mode du poste de conduite se subdivise en deux, l'un étant associé à la zone volant/compteur, l'autre étant associé à la zone pédalier/levier de vitesse.

Les facteurs « consigne » et « image » n'expliquent pas les caractéristiques des modélisations des points de fixation. La mise en évidence de stratégies d'exploration visuelle qui seraient propres à la résolution d'une consigne ou à l'exploration d'un produit n'est donc pas certaine.

6.2.2.5 Analyse comparative des modèles de densité

La distance de Kullback-Leibler entre modèles de densité est employée pour comparer deux à deux les fonctions de distribution des fixations de chaque condition expérimentale. La dimension 1 discrimine d'un côté les modèles du produit 207, et d'un autre côté les modèles du produit C6. Cependant, il apparaît que le 'stress' du modèle MDS (critère de qualité de la modalisation) n'est pas acceptable (sup à 10%) car c'est 4 dimensions qui modélisent au mieux l'espace de projection entre modèles (stress= 0,05). Une Classification Ascendante Hiérarchique est alors appliquée sur le tableau de distance de Kullback-Leibler entre modèles de densité de Parzen de chaque condition expérimentale, et met en évidence les groupes multidimensionnels.

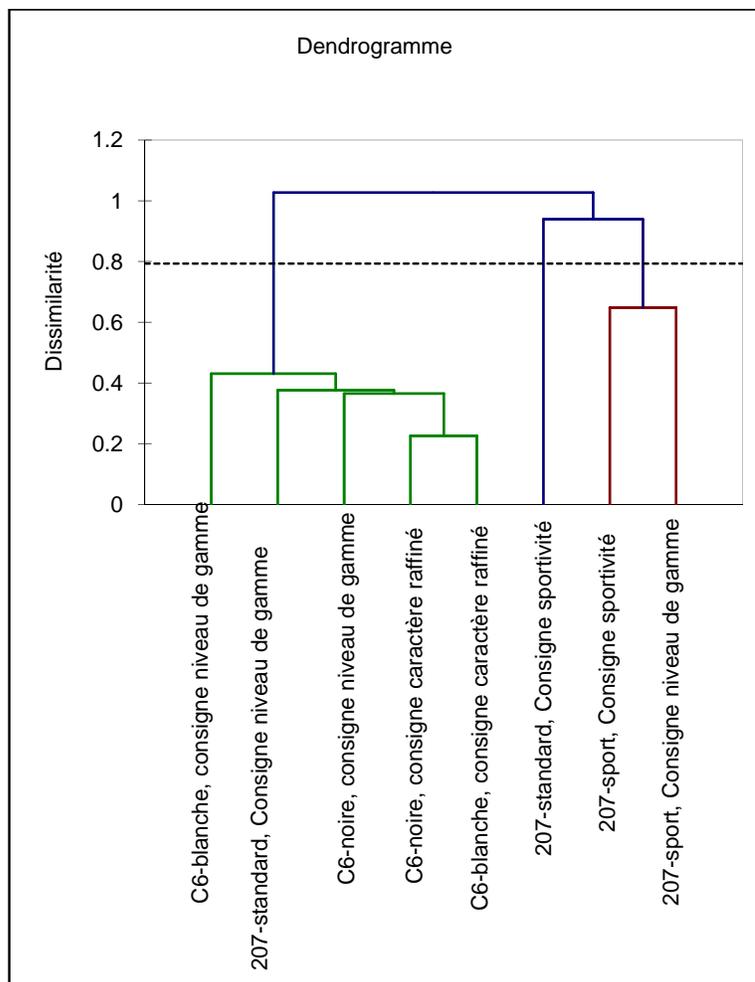


Figure 21 Dendrogramme représentant la Classification Ascendante Hiérarchique des distances entre modèles de densité

La Classification Ascendante Hiérarchique (Figure 21) crée à un premier niveau une classe contenant les images de 207 et une classe contenant les images de C6. Il existe cependant une « mauvaise attribution », avec la présence dans la classe C6 du modèle 207-standard, consigne niveau de gamme. Cette mauvaise attribution va limiter la robustesse de l'interprétation. On peut cependant proposer avec prudence que les distributions spatiales des fixations oculaires semblent plus similaires pour une même famille de produits (207-sport, 207-standard) quelle que soit la consigne qu'inter produits. Cette classification indique aussi que le déploiement de l'attention visuelle, représenté par une fonction de distribution de densité, est assez similaire pour deux modalités visuelles proches quelle que soit la consigne. Cependant, les facteurs expérimentaux n'expliquent pas clairement les propriétés de ces fonctions de distribution de densité. Cela nous amène donc à nous intéresser à la variabilité intrinsèque de chaque condition expérimentale à partir des distributions de densité.

6.2.2.6 Homogénéité des essais intra conditions expérimentales

Si la prise d'information est bien influencée par les conditions expérimentales, alors il doit exister une homogénéité de l'exploration visuelle de la scène entre essais d'une même condition. Dans ce cas

le modèle de densité va être spécifique à sa condition expérimentale, et les essais d'une condition seront bien représentés par ce modèle. Afin d'observer s'il est plus vraisemblable qu'un essai¹⁷ appartienne au modèle de densité de la condition expérimentale duquel il est issu plutôt qu'à un autre modèle, on calcul pour chaque essai la densité moyenne des zones fixées sur chaque modèle de densité. Si la densité moyenne est supérieure pour le modèle dont l'essai est issu, alors l'attribution est bonne, car cela signifie que les zones de fixations de l'essai sont fixées de manière commune avec les autres essais de la condition, sinon l'attribution est erronée. La matrice de confusion synthétise cette observation sous forme de pourcentages d'attributions correctes.

	207-sport, C1	207-sport, C2	207-standard, C1	207-standard, C2	C6-noire, C1	C6-noire, C2	C6-blanche, C1	C6-blanche, C2
207-sport, Consigne 1	73	27	0	0	0	0	0	0
207-sport, Consigne 2	0	100	0	0	0	0	0	0
207-standard, Consigne 1	0	0	100	0	0	0	0	0
207-standard, Consigne 2	14	14	0	64	0	7	0	0
C6-noire, consigne 1	0	0	0	0	73	18	0	9
C6-noire, consigne 2	0	0	0	0	0	100	0	0
C6-blanche, consigne 1	0	0	0	0	0	0	100	0
C6-blanche, consigne 2	0	0	0	0	0	9	0	91

Tableau 5 Matrice de confusion de l'appartenance des essais aux modèles de densité. En ordonnée, modèle d'origine, en abscisse, modèle d'affectation.

Consigne 1 : Caractère sportif pour 207 et caractère raffiné pour C6

Consigne 2 : caractère haut de gamme

La matrice de confusion (Tableau 5) indique que pour 4 conditions sur les 8 (207-sport consigne niveau de gamme, 207-standard consigne sportivité, C6-noire consigne niveau de gamme, C6-blanche consigne caractère raffiné) 100% des essais sont correctement attribués. Cela signifie que la variance intra modèle est faible, donc la modélisation significative de la condition expérimentale (mais pas nécessairement des facteurs expérimentaux eux-mêmes car les effets de la consigne et de l'image peuvent se combiner).

On constate que le cas « 207-standard consigne niveau de gamme » est associé à une forte confusion (64% de bonne attribution seulement). Ce cas était justement mal positionné dans la classification des distances inter modèles. On peut donc considérer que la variance interne de cette condition est forte et peut expliquer lors de la C.A.H. sa catégorisation dans le groupe C6 et non 207. Hormis ce cas, l'ensemble des autres cas pour lesquels le taux de réattribution n'est pas de 100% voient leurs essais réattribués à la même famille de produit. Il y a donc bien une similarité exploratoire intra produit, que l'on peut attribuer à des effets de la configuration de la scène. On note que pour C6, les deux conditions qui ont 100% de bonnes affectations sont opposées en termes de facteurs expérimentaux, limitant la possibilité d'attribuer ces bonnes affectations à un facteur expérimental.

6.2.2.7 Divergence essai/modèle au cours du temps

Alors que l'on a observé dans la section précédente que le comportement du regard était homogène entre essais pour certaines conditions expérimentales, on cherche à observer s'il existe des

¹⁷ Un essai est composé de l'ensemble des fixations enregistrées sur participant pendant une tâche évaluative

phases temporelles pour lesquelles les participants observent les mêmes zones. Une hypothèse sous-jacente proposée par Tatler [Tatler05] est que plus le temps passe, moins le facteur de saillance est fort, et plus les facteurs individuels sont actifs. L'homogénéité des zones fixées serait donc supérieure au début de l'exposition.

La moyenne des densités des zones de fixation est calculée par rang sur le modèle dont elles sont issues. Si l'homogénéité entre participants augmente avec le temps, alors la densité moyenne doit aussi augmenter puisque chaque fixation sera positionnée sur des zones de forte densité. La moyenne de densité par rang de fixation suit la fonction exposée Figure 22.

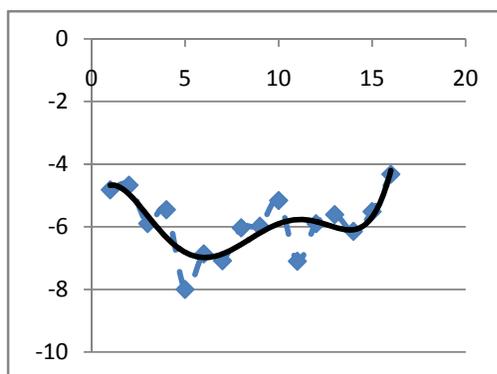


Figure 22 Evolution de la densité inter participants en fonction du rang de fixation. En pointillées les mesures, en trait plein la courbe de tendance polynomiale

L'homogénéité inter individus semble diminuer fortement pendant les 4 premières fixations, puis oscille et remonte. En termes de stratégie exploratoire cela signifie que pour les 4 premières fixations (environ la première seconde), le regard des participants ne fixe pas les mêmes zones, puis va se disperser sur l'image, puis progressivement se synchronise sur des zones communes. La densité de Parzen calculée par fenêtrage temporel montre que c'est la zone centrale qui recueille les fixations les plus tardives. Cependant, ces résultats ne sont pas suffisamment robustes pour pouvoir émettre des hypothèses sur les facteurs qui orienteraient ce processus d'homogénéité spatio-temporelle inter individuelle d'observation de l'image.

6.2.2.8 Conclusion

Les traitements qui ont été appliqués dans cette première section ont pour fonction de dégager des informations d'abord descriptives, puis comparatives, sur le comportement oculomoteur. On voit apparaître que les fixations ne se posent pas au hasard sur l'image; le type de produit (207 ou C6) semble produire des distributions de fixation assez similaires. Les consignes ne semblent pas avoir beaucoup d'effets sur le comportement du regard. A ce niveau d'observation, il n'est pas évident de mettre en évidence des stratégies de prise d'information propres à des consignes ou à des niveaux de jugements. La forte similarité des modèles de densité quels que soient les facteurs expérimentaux, et l'absence d'effets quantifiables des consignes obligent à être très prudent lors de l'interprétation de l'analyse des zones d'intérêts, qui va être présentée dans le paragraphe suivant.

6.2.3 Analyse du comportement du regard par zones d'intérêt

6.2.3.1 Méthode

L'analyse des zones d'intérêt a pour objectif de faire émerger les zones de l'images qui sont spécifiquement traitées par l'attention visuelle au cours d'une tâche, afin de tenter de relier la recherche et l'intégration d'informations avec les fixations et saccades effectuées par le regard. Pour

analyser le comportement oculomoteur par zones d'intérêt, il est nécessaire de définir les zones spatiales que l'on souhaite comparer. Nous privilégions ici une démarche d'analyse avec le minimum d'hypothèses *a priori* quant aux zones vers lesquelles s'oriente l'attention et quant aux informations visuelles pertinentes pour la consigne. A la suite des mesures, on va ainsi définir des groupes de fixations, que l'on considèrera associés à une zone d'intérêt. Le point de départ de cette analyse est la liste des barycentres des fixations obtenus par calcul des « Mean shift » et suivi d'une correction manuelle. On considère alors que ce sont ces modes qui vont être comparés. Cependant, parce que la modélisation par mélange de gaussiennes retourne un modèle optimisé en termes de position des modes et de capture des fixations uniformément réparties, cette liste de barycentre va servir à initialiser un mélange de gaussiennes dont le nombre de mode sera fixe, et dont la liberté sera d'ajouter un mode uniforme et de repositionner les modes prédéfinis.

La démarche proposée ici est d'effectuer une première étude descriptive mono variée du traitement oculaire par zones d'intérêt, puis analyser ces zones par une approche multi variée. L'objectif de l'approche multi variée effectuée par une Analyse en Composantes Principales est de caractériser de manière synthétique les zones d'intérêt grâce aux variables décrivant le comportement du regard. Puisque l'on souhaite dégager les zones qui reçoivent une attention visuelle particulière, on sélectionne plusieurs variables que l'on considère refléter l'attention visuelle : on va chercher à dégager les zones qui sont fixées le plus rapidement, le plus longuement, le plus souvent, et vérifier si ces zones peuvent s'expliquer par les conditions expérimentales. Les « individus » étudiés sont ainsi les zones d'intérêt définies *a posteriori*, et les variables sont les caractéristiques des mouvements oculaires sur chaque zone. Puisque les zones sont définies sur la base des distributions spatiales des fixations de chaque condition, on ne peut comparer directement les zones entre conditions expérimentales. Les variables qui sont sélectionnées pour les analyses mono variées sont les suivantes :

- temps cumulé par zones. Ce descripteur est établi en cumulant pour une zone les durées des fixations qui lui sont attribuées. Plus une zone est observée, plus cette variable sera élevée
- nombre d'entrées du regard par zone. Il s'agit là de la proportion de fixations attribuées à la zone par rapport au nombre total de fixations. Plus une zone est observée régulièrement (quel que soit le temps d'observation), plus la variable sera élevée
- rang de fixation. Cette variable a pour fonction de dégager si une zone est traitée plutôt rapidement ou plutôt tardivement dans la séquence. Pour chaque zone et chaque séquence est relevé le rang de la première fixation dans la zone. Cette information sera traitée par le score de Bradley-Terry, qui sera d'autant plus élevé que la zone sera fixée tôt dans la séquence par rapport aux autres zones
- moyenne des durées de fixations : cette variable a pour objectif de distinguer les zones observées longuement par peu d'individus et les zones observées rapidement par beaucoup d'individus.

Enfin, la matrice de transition est établie par conditions pour dégager si certaines liaisons entre zones sont significativement sur représentées, qui pourraient être attribués à des liens sémantiques ou perceptifs.

Pour l'analyse multi variée, en sus des variables utilisées dans l'analyse mono variée (temps cumulé par zone, nombre d'entrées par zone, rang des fixations, moyenne des durées de fixations), les indicateurs comportementaux suivants sont calculés par zones d'intérêt :

- densité moyenne. Cette variable est établie en moyennant les valeurs de densité de Parzen mesurés sur les lieux de chaque fixation attribuée à la zone. Plus la variance de

la zone sera faible, plus cette densité moyenne sera grande. Cela indiquera que le regard s'est orienté précisément sur la zone.

- moyenne des amplitudes de saccades précédant la fixation de la zone et moyenne des amplitudes de saccades suivant la fixation de la zone. Si la première variable est grande, cela signifiera alors que le regard part de loin pour arriver sur la zone, et donc que la probabilité qu'elle soit attentionnellement traitée est forte. Si la seconde variable est grande, cela peut signifier que la zone est un point de départ pour une exploration éloignée de l'image. Si ces deux variables sont faibles cela peut manifester une exploration de la scène par proche voisins. Enfin, si ces deux variables sont élevées, alors cela peut simplement manifester le fait que la zone est excentrée dans la scène mais qu'elle attire l'attention.
- Timecode¹⁸ moyen de la première fixation dans la zone. Plus la zone sera fixée tôt, plus le timecode moyen de première fixation sera faible. Cela peut manifester la force d'attractivité d'une zone (quelle que soit la raison de cette attractivité).
- Moyenne des rangs de fixations de la zone. Plus cette variable est grande, plus la probabilité que la zone soit fixée tôt est faible.

Pour l'analyse mono variée, les figures décrivant les variables sont présentées en annexe (section 1.2.1), et la synthèse des résultats par condition est présentée ci-dessous. Pour l'analyse multi variée, la projection de l'A.C.P est présentée dans le document.

6.2.3.2 Produit 207

✓ *207-sport x consigne sportivité*



Figure 23 Centre des modes: présentation des zones d'intérêt

La Figure 23 représente les centres des modes de fixations fournis par les mélanges de gaussiennes. Puisque ces modes correspondent à des regroupements de zones, ce sont les caractéristiques des fixations et saccades attribués à chacun des modes qui sont comparées pour en déduire la manière dont l'attention visuelle a traitée les différentes zones de l'image.

▪ *Analyse mono variée*

(Voir annexe 1.2.1.1). Deux zones sont principalement traitées, en fréquence et en durée : volant/compteur et haut de console centrale. La zone volant/compteur est traitée la plus tôt. Les transitions se font principalement entre les zones volant/compteur et haut de console centrale, ainsi

¹⁸ Le Timecode est le relevé temporel relatif à la séquence (nul pour le début de l'exposition et égal à 8 secondes à la fin de l'exposition)

que du pédalier vers le levier de vitesse. Les durées de fixation ne sont pas significativement différentes selon la zone ($p=0.6$). La durée moyenne des fixations est de 350ms.

▪ *Analyse multi variée*

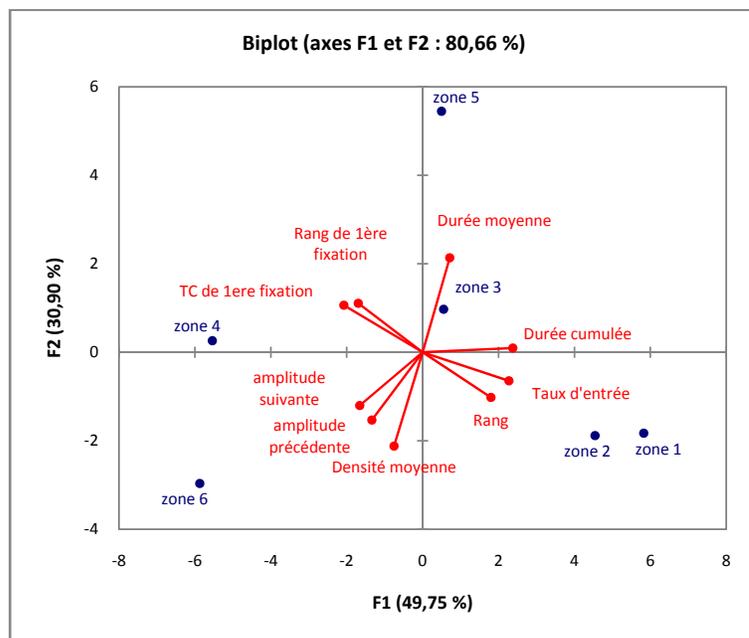


Figure 24 Représentation des zones d'intérêt sur les deux premières Composantes Principales

80% de l'inertie est exprimée dans la représentation Figure 24. On observe 3 groupes :

- zones 1 et 2 (compteur & aérateur central-centre l'image) : zones fixées tôt dans la séquence, de nombreuses fois, longuement.
- zones 3 et 5 (pédalier & levier de vitesse) : zones fixées plus tardivement que les zones 1 et 2, avec des plus petites saccades et des durées de fixations plus grandes que les zones 4 et 6,
- zones 4 et 6 (boite à gants et aérateur-extrémité droite) : zones traitées tardivement, rarement, rapidement, et avec de longues saccades.

Notons que la zone 5 est traitée avec des durées de fixations spécialement longues.

✓ *207-sport x consigne niveau de gamme*



Figure 25 Centre des zones d'intérêt

▪ *Analyse mono variée*

(Voir Annexe 1.2.1.2). On observe sur la Figure 25 que les zones sémantiques pédalier et levier de vitesse sont associés à un seul mode, indiquant qu'ils sont visuellement moins bien discriminés que pour la consigne sportivité. La zone centrale est plus traitée que les autres, tant en fréquence qu'en cumul. La zone volant/compteur est traitée la plus tôt, mais moins régulièrement que dans le cas de la sportivité. La consigne niveau de gamme pourrait donc produire une exploration moins locale, mais aussi un moins grand intérêt d'observation de la zone volant/compteur. Ici aussi, les transitions principales ont lieu entre la zone volant et la zone centrale et ce sont ces mêmes zones qui sont traitées les plus tôt. Le facteur zone n'a pas d'effets significatifs sur la durée des fixations. Enfin, à la différence du cas précédent, il y a presque autant de fixations de durée moyenne (entre 250 et 500ms) que de courtes fixations. La durée moyenne des fixations est de 420ms, donc supérieure au cas d'évaluation de la sportivité.

▪ *Analyse multi variée*

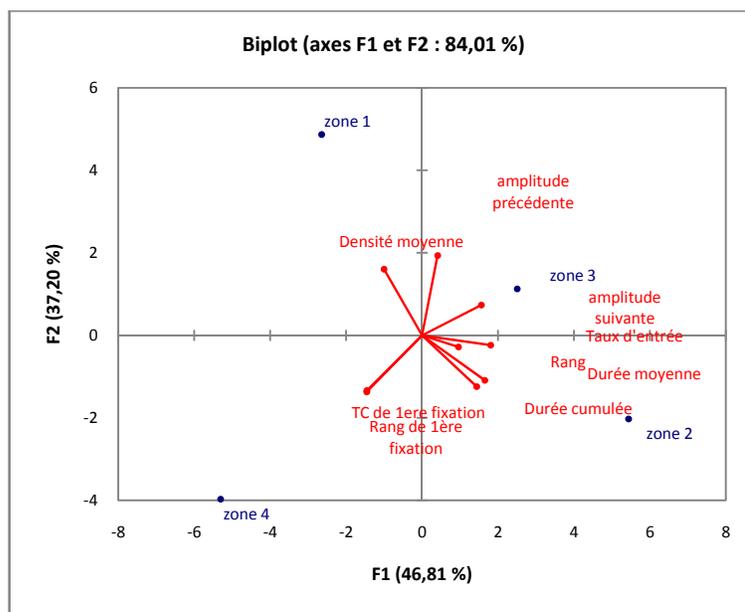


Figure 26 Représentation des zones d'intérêt sur les deux premières Composantes Principales
84% de l'inertie est exprimée dans cette représentation bidimensionnelle.

- les zones 2 (console centrale) et 3 (volant/compteur) se distinguent des zones 1 (boite à gants/ aérateur droite) et 4 (pédalier/levier de vitesse) car ces dernières sont traitées plus rarement et plus rapidement.
- la zone 1 est distinguée de la zone 4 car cette dernière est traitée plus tardivement, et avec de plus petites amplitudes de saccades précédentes
- la zone 2 se distingue de la zone 3 car la zone 3 est traitée de manière plus locale (densité moyenne plus forte), mais aussi moins longuement et moins régulièrement.

On pourrait alors poser l'hypothèse que pour la consigne niveau de gamme, les attributs de sportivité (pédalier/levier de vitesse) sont traités moins localement par le regard (un seul mode les représentant). La console centrale devient une zone privilégiée d'observation centrale de la scène (nombreuses et longues fixations), indiquant que le processus de recherche d'informations locales est peut être moins fort que pour la consigne sportivité.

✓ 207-standard x consigne sportivité



Figure 27 Centre des zones d'intérêt

▪ Analyse mono variée

(Voir Annexe 1.2.1.3). Comme dans les deux cas précédents, ce sont les zones volant/compteur et centre de console qui sont les plus traitées, en temps et en fréquence. Cependant, différemment des cas sur la 207-sport, c'est la zone centrale qui est fixée le plus tôt, suivie de la zone volant. Ici encore les principales transitions ont lieu entre ces deux zones. Comme dans le cas « 207-sport x consigne sportivité », il y a une surreprésentation des courtes fixations, tandis que la durée moyenne des fixations est de 350ms, sans qu'il n'y ait d'effets de la zone.

▪ Analyse multi variée

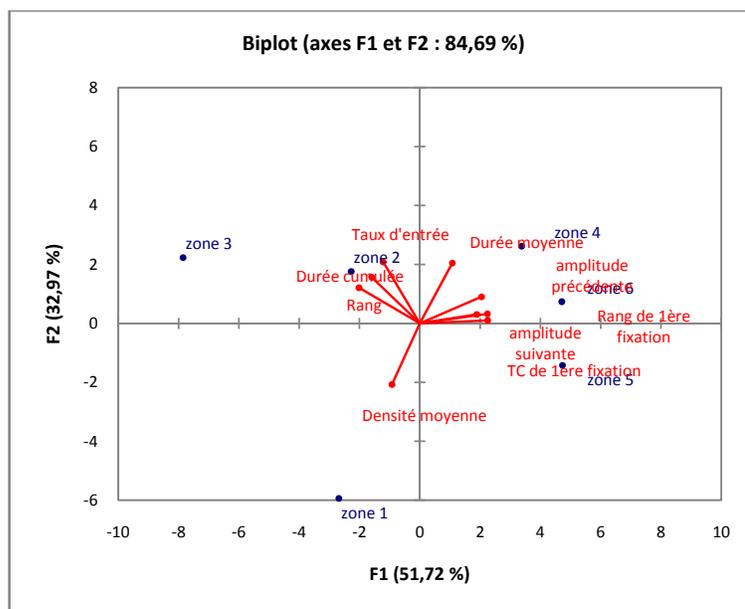


Figure 28 Représentation des zones d'intérêt sur les deux premières Composantes Principales

85% de l'inertie est exprimée par cette représentation.

- Les zones 2 et 3 (volant, aérateur central), qui sont traitées les plus souvent, les plus longuement et les plus tôt.
- la zone 1 (rétroviseur central) des autres zones, étant fixée avec des fixations plus courtes et sur un espace de faible densité

- Les zones 4 (levier de vitesse), 5 (pédalier) et 6 (aérateur droite) sont traitées plus tardivement et avec de plus grandes amplitudes de saccades. La zone 4 est traitée plus souvent que les zones 5 et 6, et sur des densités plus faibles.

Ainsi, la zone volant et aérateur centrale sont les zones vers lesquelles l'attention revient le plus souvent. Le rétroviseur est traité rapidement et peu de fois, tandis que la zone levier de vitesse est traitée avec des fixations de longue durée.

✓ *207-standard x consigne niveau de gamme*



Figure 29 Centre des zones d'intérêt

▪ *Analyse mono variée*

(Voir Annexe 1.2.1.4). Les modes sont positionnés de la même manière que pour cette même image pour la consigne sportivité, mais le mode sur le rétroviseur a disparu. On observe ici l'émergence de 3 zones : zone centrale supérieure, zone centrale levier de vitesse, et zone volant/compteur. En temps cumulé, c'est la zone centrale supérieure qui gagne tandis qu'en fréquence, c'est la zone levier de vitesse. C'est la zone centrale supérieure qui est traitée le plus tôt, comme dans le cas précédent, suivi de la zone volant/compteur et de la zone levier de vitesse. Les transitions principales ont lieu entre la zone centrale supérieure et la zone volant, mais les transitions avec la zone levier de vitesse sont aussi surreprésentées. Encore plus que dans tous les cas précédents, les fixations de faible durée sont surreprésentées. A la différence de tous les cas précédents, il y a un effet du facteur zone sur les durées de fixation, dont la moyenne est de 350ms, et où la zone pédalier et levier de vitesse sont traités avec des fixations significativement plus longues que les autres zones.

▪ *Analyse multi variée*

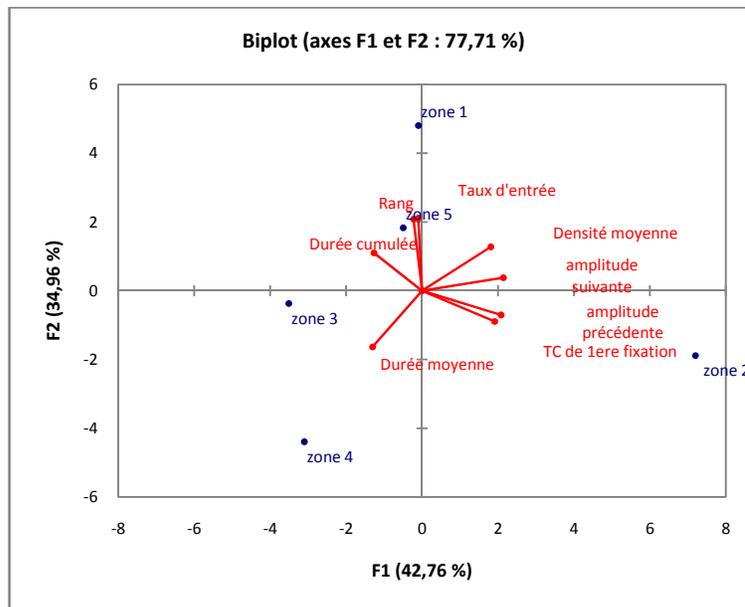


Figure 30 Représentation des zones d'intérêt sur les deux premières Composantes Principales

78% de l'inertie est exprimée dans cette représentation bi dimensionnelle.

- la zone 2 (boîte à gants) est la moins traitée, et la plus tardivement
- la zone 1 (aérateur central) est la zone fixée le plus souvent et le plus tôt, suivie par la zone 5 (volant)
- la zone 4 (pédalier) possède les plus faibles durées de fixation
- la zone 3 (levier vitesse) est celle qui est la plus observée au cumul.

Ainsi, la zone centrale et le volant sont rapidement et souvent traitées, La zone pédalier est observée avec des courtes durées de fixations. Cela peut s'expliquer car il n'y a pas d'informations visuelles pour y retenir le regard, qui se serait dirigé sur cette zone par un processus de recherche d'information. Enfin, la zone de droite de l'image est explorée tard dans la séquence, une fois que les autres zones ont été investiguées. Le participant n'y cherche donc pas d'informations spécifiques pour la consigne.

✓ Synthèse 207

La 207-sport semble bénéficier d'un traitement visuel fort sur la zone volant/compteur, fixée plus tôt que les autres. Le cas « 207-sport x consigne niveau de gamme » possède des durées de fixations spécifiquement longues, le cas « 207-standard x consigne niveau de gamme » montre un effet significatif du facteur zone sur les durées de fixation. Pour le cas « 207-sport x consigne sportivité », on note la présence de modes localisés sur des zones d'intérêts bien discriminées : pédalier chrome, compteur blanc, levier de vitesse chrome, contour de console centrale carbone. Cependant ces caractéristiques visuelles ne s'expliquent pas par les conditions expérimentales. On remarque cependant que lorsque la consigne (sportivité) renvoie à des signes visuels spécifiques, ceux-ci semblent être fixés localement.

L'Analyse par Composante Principale permet de dégager les zones bénéficiant d'un traitement visuel particulier. Il n'apparaît cependant pas, dans ces cas précis, que l'on puisse expliquer par les conditions expérimentales (modalité visuelle et consigne) les groupes de zones décrits par l'ACP. Enfin 2 principaux modes assez similaires pour ces quatre conditions : volant/compteur et console centrale,

qui sont donc des zones traitées sans liens avec la consigne ou parce que les deux consignes y dirigerait le regard.

En termes de méthode, les informations mono variées convergent avec les analyses multi variées, mais permettent plus difficilement de dégager des groupes de zones par le traitement oculaire dont elles bénéficient. On observe que les zones visuelles sont bien discriminées spatialement et par les variables choisies, mais qu'il n'est pas aisé de lier condition expérimentale et positionnement multidimensionnel.

6.2.3.3 Etude pour le produit C6

✓ *C6-noire x consigne caractère raffiné*



Figure 31 Centre des zones d'intérêt

- *Analyse mono variée*

(Voir Annexe 1.2.1.5). La zone centrale de l'image qui englobe la console centrale et le levier de vitesse est celle sur laquelle le regard se pose le plus longuement. En fréquence, la zone volant, la zone écran et la zone centrale sont presque autant traitées. Cela signifie que les zones volant et écran recueillent des fixations plus courtes que la zone centrale, ou que cette dernière zone est une zone de stabilité sur laquelle le regard se pose longuement. La zone volant est celle qui est traitée le plus tôt, suivie par la zone écran puis la zone centrale. Les transitions entre la zone volant et la zone centrale sont surreprésentées. Il n'y a pas d'effet du facteur zone sur les durées de fixations, dont la moyenne est de 360ms pour chacune sauf pour la zone centrale dont la moyenne est à 500ms. On note cependant une surreprésentation des fixations inférieures à 250ms.

- *Analyse multi variée*

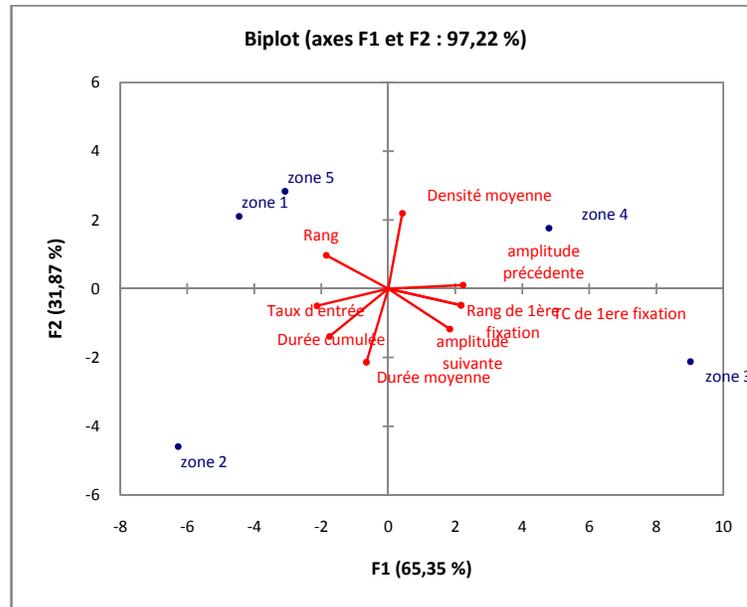


Figure 32 Représentation des zones d'intérêt sur les deux premières Composantes Principales 97% de l'inertie est représentée.

- la zone 2 (zone centrale) se distingue des autres zones par une plus grande durée moyenne des fixations, ainsi que par un long temps cumulé d'observation
- Les zones 1 (volant) et 5 (ordinateur-centre de l'image) se distinguent par un traitement tôt
- Les zones 3 (boite à gants) et 4 (haut de planche de bord-extrémité droite) apparaissent comme moins traitées en termes de nombre d'entrées, ainsi que plus tardivement.

On peut donc considérer que la zone 2 (centre de l'habitacle-centre de la console centrale) est une zone d'équilibre visuel, sur laquelle le regard reste stationné longuement, et sur laquelle il revient souvent. Les zones 1 et 5 sont des zones attractives, traitées tôt dans l'exposition, et régulièrement. Enfin les zones 3 et 4 apparaissent traitées de manière secondaire, tardivement et moins fréquemment.

✓ Cas C6-noire x consigne caractère niveau de gamme

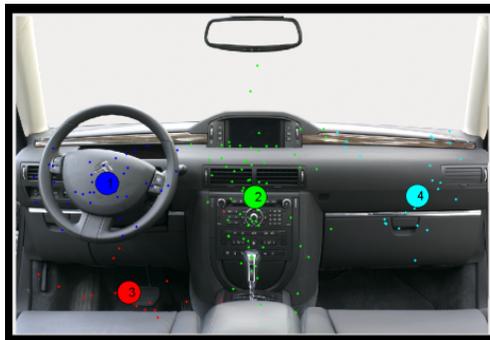


Figure 33 Centre des zones d'intérêt

- Analyse mono variée

(Voir Annexe 1.2.1.6). On note tout d'abord, par rapport à l'analyse du même produit mais sous la consigne « caractère raffiné », l'apparition d'un mode au niveau du pédalier et de la disparition des modes sur l'écran et sur la texture bois du haut de la planche de bord. Deux zones sont spécialement plus traitées visuellement : la zone volant/logo et la zone centrale/levier de vitesse, tant en temps cumulé qu'en nombre d'entrées. Ces deux zones sont traitées aussi tôt l'une que l'autre et les fixations qui les lient directement (transitions) sont sur représentées. Il y a un effet significatif de la zone sur la durée des fixations, la zone centrale étant fixée en moyenne pendant 500ms, la zone volant pendant 420ms, la zone boîte à gants pendant 380ms, et la zone pédalier pendant 200ms. La durée élevée des fixations de la zone centrale pourrait s'expliquer par l'existence d'une zone d'équilibre et la faible durée de fixation de la zone pédalier peut être expliquée par la simplicité/disparité des informations visuelles à extraire. Le regard pourrait donc s'y placer par motivation endogène, et ne pas s'y stabiliser pour des raisons perceptives.

▪ *Analyse multi variée*

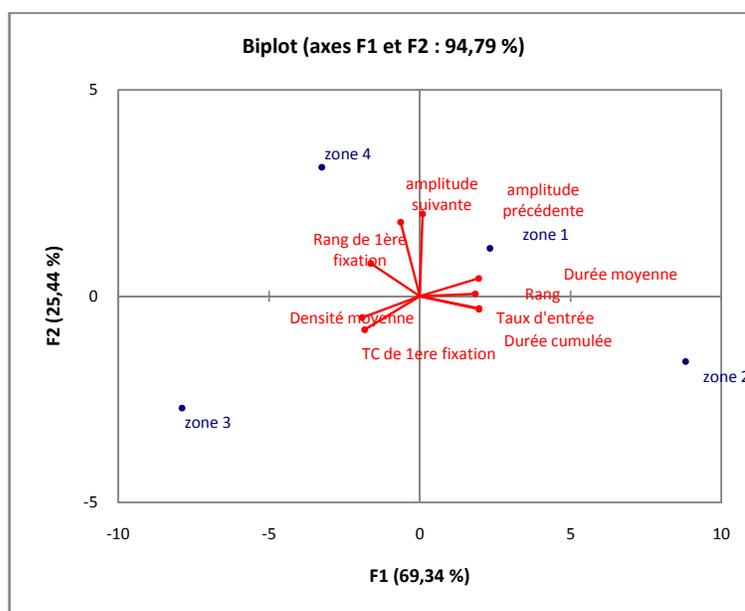


Figure 34 Représentation des zones d'intérêt sur les deux premières Composantes Principales
95% de l'inertie est représentée.

- la zone 3 (pédalier) se dégage des autres zones car étant traitée pour la première fois plus tardivement et contenant une faible densité attentionnelle
- la zone 4 (boîte à gants), puis les zones 1 (volant) et 2 (centre console centrale), sont positionnées dans un ordre croissant de temps cumulé, de fréquence de traitement
- La zone 2 se dégage par son traitement très tôt dans la séquence.

✓ Cas C6-blanche x consigne caractère raffiné

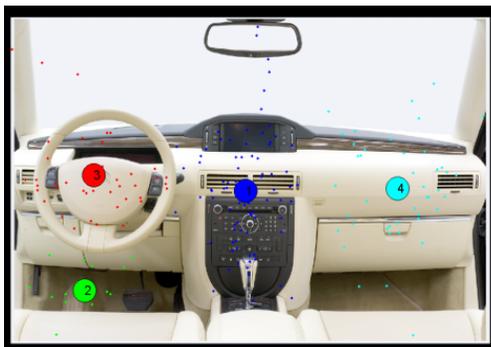


Figure 35 Centre des zones d'intérêt

▪ Analyse mono variée

(Voir Annexe 1.2.1.7). On observe que les modes sont positionnés de manière très similaire au cas précédent, qui n'a pourtant en commun ni la modalité visuelle ni la consigne. La zone centrale est spécifiquement sur traitée visuellement tant en temps cumulé qu'en fréquence. Puis à égalité viennent la zone volant et la zone boîte à gants pour le temps cumulé, tandis que la zone volant est traitée plus fréquemment que la zone boîte à gants. Les fixations sur le volant sont donc plutôt de courte durée. Chronologiquement, la première zone impactée est la zone centrale, suivit par le volant. Les transitions entre volant et centre sont surreprésentées. Il n'y a pas d'effet significatif du facteur zone sur les durées de fixation, d'en moyenne de 350ms, où les petites durées sont sur représentées.

▪ Analyse multi variée

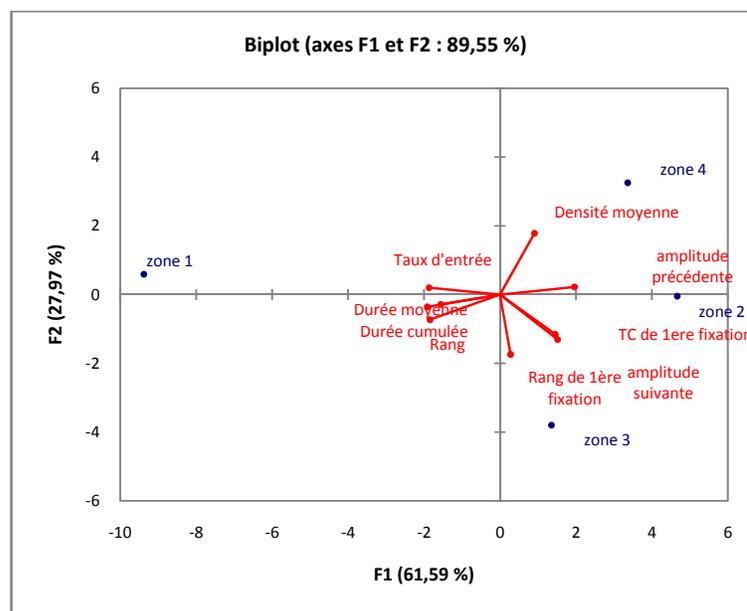


Figure 36 Représentation des zones d'intérêt sur les deux premières Composantes Principales.

90% de l'inertie est exprimée.

La zone centrale '1' se dégage des autres car étant significativement traitée plus longuement, avec un temps cumulé élevé, un grand nombre d'entrées. Elle est traitée tôt dans la séquence.

✓ «C6-blanche x consigne caractère niveau de gamme



Figure 37 Centre des zones d'intérêt

▪ *Analyse mono variée*

(Voir Annexe 1.2.1.8). Contrairement au produit « C6-blanche » sous la consigne « raffinée », le mode sur le pédalier n'est pas présent. Cela était aussi le cas pour C6-noire consigne « raffiné » avec lequel ce cas n'a en commun ni la modalité visuelle ni la consigne. De même que ce dernier cas, il existe un mode sur l'écran. Le mode écran, le mode volant, et le mode central sont traités sans que l'un soit favorisé sur l'autre. La zone volant est fixée significativement plus tôt que les autres, suivie de la zone centrale puis de la zone écran. Il n'y a pas d'effet du facteur zone sur les durées de fixation (moyenne de 380ms) et les fixations de faible durée sont surreprésentées.

▪ *Analyse multi variée*

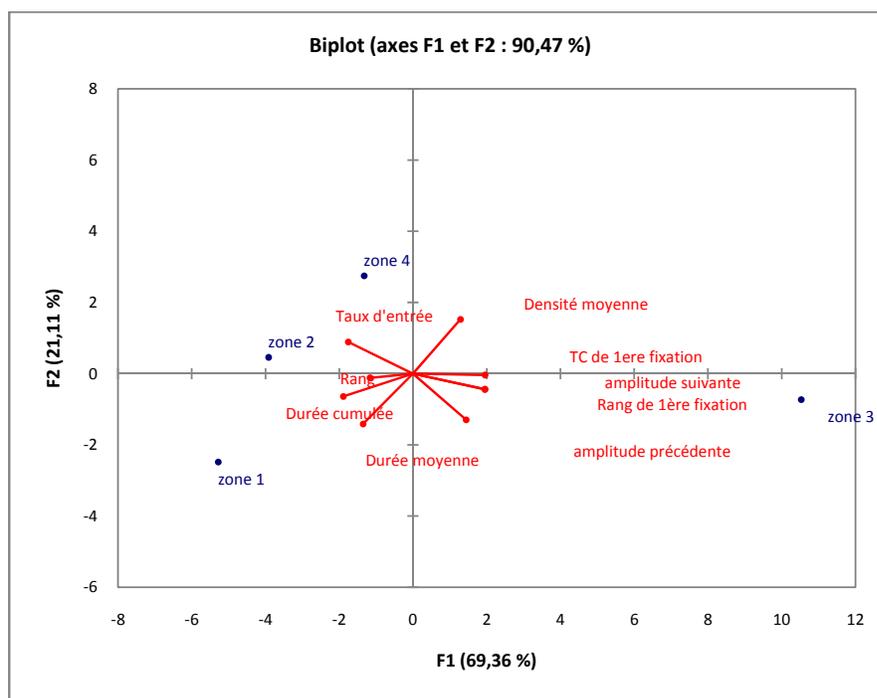


Figure 38 Représentation des zones d'intérêt sur les deux premières Composantes Principales

90% de l'inertie est représentée dans cette projection bidimensionnelle.

- o la zone 3 (boîte à gants) est traitée la moins intensivement et la plus tardivement

- La zone 4 (ordinateur-centre de l'image) est traitée de manière plutôt localisée (densité forte), on peut donc considérer que le regard s'y est spécifiquement posé
- La zone 1 (centre de la console centrale) se dégage par des durées de fixation élevées, et un temps cumulé élevé.

Il n'y a pas de zone qui soit traitée spécifiquement plus tôt que les autres ou plus fréquemment que d'autres.

✓ *Synthèse produit C6*

Hormis dans le cas « C6-noire x consigne caractère niveau de gamme », trois modes sont fortement traités : volant, console centrale, puis soit écran, soit boîte à gants. Tant en terme de première zone fixée que de position des modes, il y a de nombreux points communs entre le cas C6-noire consigne caractère raffiné et C6-blanche consigne niveau de gamme. On ne peut donc pas considérer que les facteurs expérimentaux expliquent le traitement visuel de chaque condition. Le mode central apparaît ici comme étant très fort. Pour tous ces cas, la zone centrale semble être un point de vue privilégié depuis et vers lequel le regard fait des allers et retours avec les autres zones. La zone de droite (passager) est traitée plus tardivement dans la séquence. La zone volant est toujours une zone bien traitée sans qu'elle ne se dégage spécifiquement dans aucun des cas. Ici non plus, on ne met pas en évidence d'effets propres à une modalité visuelle ou à une consigne par l'orientation du regard sur des zones d'intérêt.

6.2.3.4 Conclusion

Le regard semble guidé par la configuration spatiale de la modalité visuelle indépendamment de la consigne ; il est ainsi difficile de relier l'attention visuelle à des spécificités de la consigne. Parfois le volant est traité en premier, parfois c'est la zone centrale. Lorsqu'il existe des attributs visuels répondant spécifiquement à la consigne, ces derniers bénéficient d'un traitement visuel localisé (consigne sportivité / 207-sport). Si ces variables décrivent les caractéristiques de l'orientation du regard sur les zones d'intérêts, elle ne permet pas dans le cadre de ce paradigme de dégager des processus de recherche d'informations spécifiques.

L'analyse multi variée par composante principale, sans contredire les résultats établis à partir des variables dans la phase mono variée, permet d'aller plus loin en identifiant les variables qui se corrélient, en regroupant les zones en fonction des variables oculomotrices, et en dégageant les variables qui les discriminent. Ce traitement permet de discriminer les zones selon les spécificités du traitement visuel dont elles bénéficient. Une zone traitée avec une forte attention peut être mise en évidence par rapport à une zone peu traitée par le regard. Dans certaines situations on peut interpréter les raisons de cette attention particulière : s'agit-il d'un centre de gravité ? Est-ce une zone qui surprend ? Une zone informative ? Une zone centrale ? De la même manière ce traitement peut permettre de regrouper les zones, et donc de dégager des niveaux d'attractivité ; par exemple la zone volant et la zone centrale sont fixées longuement et tôt, mais pas nécessairement pour des raisons de traitement d'informations visuelles propres à la consigne ; la zone pédalier et la zone levier de vitesse sont fixées plus tardivement mais avec des densités élevées et régulièrement : cela manifesterait-il un intérêt attentionnel spécifique pour ces zones ?

De manière déductive, en partant d'hypothèses sur l'importance d'une zone ou sur la similarité entre deux zones, on peut observer si effectivement une zone bénéficie du traitement attentionnel prédit, si elle se dégage bien des autres zones, ou encore quelles dimensions distinguent deux zones.

Après avoir analysé en détail le comportement oculomoteur par découpage de la scène visuelle en zones, et nous être rendu compte de la difficulté à dégager les processus attentionnels activés du fait à la fois de la similarité macroscopique des distributions de fixation et à la fois de la variabilité pour

chaque condition expérimentale, des comportements oculomoteurs, nous allons chercher dans la section suivante à observer s'il existe des groupes de stratégies de prise d'information visuelle.

6.2.4 Classification des stratégies oculomotrices

Les parties précédentes ont exposées des résultats permettant de mieux comprendre la manière dont s'oriente le regard sur la scène, et les zones visuelles qui se distinguent en termes de traitement visuel, par une approche combinée (tous les essais de chaque condition expérimentale sont regroupés). Dans cette section, nous exposons les résultats permettant de vérifier si différentes stratégies oculomotrices peuvent se caractériser, et le cas échéant si les différentes stratégies peuvent s'expliquer par les facteurs expérimentaux par exemple. La première observation est de nature mono variée spatiale et la seconde observation est de nature multi variée

6.2.4.1 Stratégies spatiales caractérisées par la distance de Mannan

La distance de Mannan est une métrique permettant de quantifier la similarité spatiale entre deux nuages de points au nombre d'individus différents. Ainsi deux distributions de fixations peuvent être comparées et leur distance quantifiée. Plus la distance est grande, moins les individus sont positionnés de manière homogène. On calcule ainsi la distance de Mannan entre tous les essais mesurés dans l'expérimentation monadique. On effectue une Classification Ascendante Hiérarchique sur le tableau de dissimilarité inter essais (annexe 1.3), qui dégage 3 groupes d'individus. Les groupes sont déséquilibrés car contenant 71, 9 et 10 individus. La constitution de ces groupes n'est pas liée à un type de consigne observé ou à un type d'image. On ne peut donc pas affirmer qu'il existe des classes de stratégies qui seraient liées à la recherche d'informations visuelles spécifiques et donc similaires par l'étendue spatiale des distributions des fixations.

6.2.4.2 Caractérisation des similarités de stratégies visuelles par Analyse par Composante Principale

Une analyse multi variée des caractéristiques du comportement oculomoteur a pour objectif de représenter de manière optimale la distance multidimensionnelle entre essais sur la base du comportement du regard. Les variables qui vont caractériser les essais sont les suivantes : 1/ nombre de zones traitées dans la séquence (reflète la motivation à explorer l'image), 2/ densité moyenne (indique si le participant fixe des zones locales communes avec les autres participants de la condition expérimentale), 3/ moyenne des durées des fixations (caractérisation de la manière dont l'image est explorée temporellement, lié aux stratégies d'exploration ambiante *versus* locale), 4/ moyenne des amplitudes de saccades (indiquant elle aussi le type de stratégies exploratoire, mais cette fois spatialement), et 5/ nombre de fixations dans la séquence (reflète une exploration locale si faible et globale si élevée). Ces variables peuvent décrire une stratégie de prise d'information, mais aussi un comportement général ou une manière de résoudre la consigne sans lien avec les informations visuelles intégrées pour la tâche. Les « individus » pour lesquels ces variables sont quantifiés sont les essais (séquence de fixations par un participant).

Afin de dégager l'existence de stratégies de groupes d'essais de toutes les conditions expérimentales sur la base du comportement visuel, une Analyse par Composante Principale est effectuée.

intégrer directement pour le jugement les informations qu'elle contient localement ? On ne peut que réfuter cette affirmation car c'est peut être un processus de saillance visuelle qui a attiré l'attention, et qui a permis son encodage mnésique. *A posteriori* cette zone est aisément mémorisée et attribuée par erreur à un attribut jugé.

Il apparaît pour la modalité visuelle 207-sport, pour la consigne sportivité, que 85% des sujets évoquent le pédalier, le levier de vitesse et les compteurs. 60% des sujets évoquent le contour carbone et les sièges en cuir, et 30% des sujets évoquent l'instrumentation. Pour la consigne niveau de gamme, 80% des sujets évoquent le levier de vitesse et les compteurs, 65% évoquent les sièges en cuir et 40% le pédalier.

Pour la modalité 207-standard, pour la consigne sportivité, 65% des sujets évoquent les compteurs, 55% évoquent le pédalier, et 40% des sujets évoquent le levier de vitesse. Pour la consigne niveau de gamme, 70% des sujets évoquent les matériaux des sièges et du contour de console centrale, 60% évoquent le levier de vitesse, et 40% évoquent le compteur.

Pour les modalités C6, que ce soit C6-noire ou C6-blanche, pour la consigne raffinement, 95% des sujets évoquent la couleur globale et 85% évoquent les matériaux : planche de bord, sièges. Seuls 35% des sujets évoquent l'écran. Pour la consigne niveau de gamme, 100% des sujets évoquent la couleur globale, et 80% les matériaux.

On constate que pour C6, le choix ne semble pas s'effectuer sur des critères visuels locaux mais plutôt sur la globalité de l'image, sur la base de textures et couleurs. Cela explique ainsi l'existence pour ce produit d'une exploration moins locale que pour 207.

Globalement, le nombre de zones citées est limité. Il est habituellement admis que 4 items peuvent être stockés simultanément lors d'une tâche non mnésique. Ce faible nombre d'éléments évoqués vient donc peut être de contraintes mnésiques. Cela peut aussi provenir du fait que les éléments cités sont ceux qui définissent le mieux la catégorie sémantique « voiture ». Notons que l'on n'observe pas de corrélations entre le taux de citation des zones et le temps d'observation par zone, mais pas d'incohérences non plus. L'étude des verbalisations permet de confirmer le fait que des éléments visuels locaux sont intégrés dans le processus évaluatif pour 207, mais que pour C6 ce sont plutôt des impressions globales. Du fait de la complexité des liens entre la position du regard, le jugement et la reconstruction verbale *a posteriori* du jugement, on n'observe pas d'effets évidents entre la sélection visuelle de zones, l'évaluation, et la verbalisation.

6.2.5 Conclusion

Notre étude sur le comportement du regard en situation d'évaluation monadique a débuté par l'observation des notes attribuées à chaque version des deux produits. On observe que les deux versions de chaque produit sont évaluées différemment, sans incohérences avec les intentions des concepteurs. On souhaite alors étudier si la prise d'information visuelle est liée à la consigne, à l'image, ou à un facteur idiosyncrasique. Une analyse globale par densités des distributions des fixations est effectuée, montrant que les distributions sont très similaires entre les différentes conditions expérimentales. On met en évidence des tendances temporelle et spatiales dans le traitement visuel des scènes, sans que l'on puisse dégager un réel effet de la consigne, donc de la demande d'informations pour l'évaluation. L'étude des zones d'intérêt permet de dégager des éléments visuels vers lesquels s'oriente le regard, et de les discriminer selon leur capacité à attirer l'attention. On n'identifie cependant pas clairement des zones spécifiquement fixées pour une consigne. L'étude des verbalisations permet d'observer que lorsque les éléments verbalisés sont locaux et distincts, alors le traitement visuel est aussi plutôt local. Cela ne lie pas prise d'information et jugement, mais permet de prédire la qualité du lien entre la mesure oculométrique et le traitement locale des informations visuelles. En effet, si le participant est dans l'incapacité à dégager verbalement

des attributs visuels, alors cela peut prédire que l'attention visuelle ne s'est pas orientée vers des éléments locaux (sans qu'il n'y ait de lien entre la zone fixée et le jugement).

Ce paradigme d'évaluation monadique et ces méthodes d'analyse montrent qu'il est difficile de mettre en évidence les zones vers lesquelles s'oriente le regard lors de l'évaluation. Plusieurs raisons peuvent expliquer cela :

- prise d'information optimisée spatialement pour limiter les mouvements du regard
- consigne ne requérant pas de fixer des informations locales
- éléments visuels nécessaires à la consigne absents de l'image
- multiplicité des facteurs qui guident l'attention (mnésiques, physiques, etc.)
- comportement visuel fortement idiosyncrasique pour ces conditions
- dissociation entre la tâche évaluative et le comportement de recueil d'informations visuelles
- biais d'observation centrale

Le paradigme d'évaluation monadique d'habitacles de véhicules ne donc semble pas mettre en évidence de relations entre zones d'intérêts et demande de tâche, ni de groupes de stratégies de prise d'information. Sur le type de stimuli employés, le regard serait plutôt guidé par la configuration spatiale de la scène indépendamment des demandes de la consigne. Il est donc pertinent de faire évoluer notre paradigme, et c'est à l'évaluation par paire auquel nous nous intéressons. Cette étude fait l'objet du chapitre suivant.

Chapitre 7

Expérimentation 2 : paradigme d'évaluation par paire

7.1 Matériel et méthodes

7.1.1 Objectifs

Nous avons pu voir dans la partie précédente les limites du paradigme de présentation monadique. C'est pourquoi nous nous intéressons dans cette partie à l'évaluation par paire. En effet, la présentation par paire implique que l'évaluation se fasse selon un processus de comparaison. Le regard est donc incité à s'orienter vers les zones à comparer. Si le jugement se base sur des critères locaux de décision, ces zones peuvent être spécifiquement traitées afin de résoudre la consigne. Un second intérêt de la comparaison par paire est d'éliminer le biais de centralité, car le regard doit se positionner d'un côté ou de l'autre de la scène pour explorer chaque élément de la paire en détail. Enfin, l'exposition par paire permet de comparer deux scènes contenant des informations visuelles identiques mais organisées différemment en terme spatial (produit 'A' à gauche – produit 'B' à droite et inversement des positions). Si, pour un contenu sémantique donné et une consigne identique, le traitement visuel n'est pas orienté vers les mêmes zones, alors il n'est pas possible de corréler évaluation et prise d'information locale.

Un paradigme de comparaison par paire est ainsi mis en place afin:

- De contraindre le regard à se positionner sur des éléments visuels locaux
- de limiter la prise de décision basée sur des éléments mnésiques en favorisant une évaluation basée sur les percepts
- de dégager des phases cognitives selon l'activité visuelle (comparaison des deux éléments, investigation d'un seul élément en détail,...)
- de mettre en évidence des zones visuelles spécifiquement traitées par comparaison
- de confirmer les intentions de jugement attendues par les concepteurs, et les zones d'intérêt associées
- d'approfondir les méthodes de traitements de données permettant d'analyser le comportement visuel

7.1.2 Matériel

Les stimuli utilisées sont identiques à ceux employés dans l'expérimentation monadique, mais présentés sous forme de paire : « Paire 207-standard / 207-sport », « Paire C6-noire / C6-blanche ».

Afin de contrebalancer le biais de latéralisation¹⁹ et l'effet d'ordre, deux paires sont constituées avec les mêmes images « Paire 207-standard / 207-sport » & « Paire 207-sport / 207-standard », « Paire C6-noire / C6-blanche » & « Paire C6-blanche / C6-noire ». Rappelons que le produit «Peugeot 207 » est utilisé sous deux modalités visuelles, l'une définie par les concepteurs comme étant un modèle sport, et l'autre étant le modèle standard. Les différences entre ces deux modalités sont d'une part des variations locales des matériaux représentés, et d'autre part des absences/présences de certains objets. Le deuxième produit choisi est le modèle Citroën C6, dont les différences entre les deux déclinaisons ne sont pas locales, mais plutôt ambiantes, en termes de couleur et textures ; la configuration et les objets de scène sont identiques (Planche 2).

Afin de pouvoir comparer les résultats avec l'expérience monadique, et parce que ce paradigme doit tester la reconnaissance du typage prédit par les concepteurs, les consignes choisies sont les suivantes. Pour le produit 207, dont la modalité visuelle est soit « sport », soit « standard », la première consigne est d'évaluer lequel des éléments de la paire est le plus sportif. Si les intentions des concepteurs sont bien reconnues, les deux modalités ne doivent pas obtenir les mêmes notes. La deuxième consigne est de définir l'élément le plus « haut de gamme ». Pour le produit C6, dont le typage est soit noir, soit blanc, la première consigne est d'évaluer laquelle des modalités est la plus raffinée, la seconde consigne est d'évaluer laquelle des modalités est la plus haut de gamme. Ces deux consignes se basent sur des critères ambiants, de la même manière que les deux modèles diffèrent sur des critères visuels ambiants (couleurs et textures). Des paires de photographies à haute résolution et forte dynamique de luminance sont construites, en adoptant un point de vue et une focale strictement identique pour chaque modalité visuelle.

¹⁹ Il est nécessaire de prendre en compte le fait qu'un habitacle n'est pas symétrique, c'est à dire que le poste de conduite est à gauche de chaque image. Ainsi, si le regard est fixé initialement au centre de l'image, il est beaucoup plus près du poste de conduite du produit de droite que de celui de gauche. De plus, on lit de gauche à droite ; cela peut l'ordre d'observation des images.



Planche 2 Paires composées à partir des produits 207 et C6

Alors l'exposition monadique durait 8 secondes, l'exposition par paire se fait pendant 12 secondes. Ce choix est la conséquence d'un pré test dans lequel 8 secondes s'est avéré trop court. En effet par rapport à l'évaluation monadique la quantité d'informations à traiter est deux fois plus élevée, et l'échelle est réduite par 2 (échelle 0,4). Enfin, 16 secondes apparaissent trop longues, les participants ayant fini la tâche bien avant la fin de l'exposition. On choisit donc une durée intermédiaire.

Il est expliqué au participant qu'ils vont juger des habitacles de véhicules, que la durée d'exposition est de douze secondes, et qu'ils devront ensuite justifier leur réponse. La consigne est : « vous allez observer une paire d'habitacle de véhicule. Vous devez bouger votre regard le plus rapidement possible vers l'habitacle qui vous semble le plus « sportif » (ou « haut de gamme » ou « raffiné »)

La séquence est composée d'une diapositive contenant un cercle de 5° d'angle visuel de large s'affichant au centre de l'image sur fond noir pendant une seconde, puis une diapositive contenant une image de bruit blanc pendant une demi-seconde. La diapositive contenant la paire est alors exposée pendant 12 secondes. Les participants donnent ensuite leur réponse de choix (élément de gauche ou élément de droite) puis leur évaluation de chaque élément en déplaçant un curseur sur une échelle en 9 points de pas du tout [...] à très [...]. Enfin, ils verbalisent à haute voix les zones qui ont aidé à leur décision et en les désignant sur un tirage papier.

Chaque participant effectue une seule évaluation et il y a au moins 10 participants par condition expérimentale. Les mêmes conditions de recrutement que le test monadique sont appliquées. Les participants sont recrutés de telle sorte d'obtenir une répartition homogène dans chaque groupe en termes d'âge dans la fourchette 20-65 ans et de sexe. Ils doivent avoir acheté une voiture neuve du segment B (bas) dans les deux dernières années, ne pas travailler dans le domaine de l'automobile ou du design, et ne pas avoir fait de test clientèle dans la dernière année. 80 essais sont ainsi effectués. Les participants sont rémunérés pour le test 40 euros).

L'exposition se fait dans une salle obscure, sur un écran de vidéo-projection à une résolution de 1400*1050, soit 125*167cm, à une distance de 2 mètres (40°*34° du champ visuel). Les mouvements oculaires sont enregistrés avec un Facelab® 4.1, c'est-à-dire un eye tracker déporté et non à casque, en mode précision avec l'aide des émetteurs infrarouges.

7.2 Résultats

Tel que cela a été présenté dans le chapitre méthodologique et qui a été ensuite appliqué dans l'étude de l'évaluation monadique, nous allons commencer par observer les notes du jugement, puis effectuerons une observation des densités, de manière descriptive et comparative. Nous nous intéresserons alors aux zones d'intérêt, et enfin à la mise en évidence de stratégies idiosyncrasiques.

7.2.1 Jugement des produits

L'objectif de cette partie est d'observer si le choix est homogène entre participants, et si les produits sont discriminés par leur note.

✓ *Produit 207*

Le Tableau 6 présente la fréquence de choix de chaque modalité visuelle (chaque version du produit) par consigne.

produit / consigne	207-sport	207-standard
La plus sportive	90% (20/22)	10 % (2/22)
La plus haut de gamme	85% (16/19)	15% (3/19)

Tableau 6 Modalité visuelle la plus fréquemment choisie

La loi binomiale indique que la 207-sport est significativement choisie comme étant la plus sportive ($B(19,22,0.5) < 0.05$) ainsi que la plus haut de gamme ($B(16,19,0.5) < 0.05$).

Le tableau suivant expose la moyenne des notes par habitacle sur 100.

produit / consigne	207-sport	207-standard
La plus sportive	70	15
La plus haut de gamme	66	30

Tableau 7 Moyenne des notes sur 100 pour chaque modalité visuelle

Un test de Student non appareillés compare les notes obtenues sur la modalité visuelle 207-sport (tab. 7-2) et dégage que cet habitacle n'est pas jugé différemment selon la consigne ($\alpha > 0,05$ – il n'est pas jugé plus sportif que haut de gamme). Le résultat est le même pour 207-standard. Un second test de Student, appareillé cette fois (car les participants qui fournissent les deux notes sont les mêmes) montre que les deux habitacles sont jugés différemment quelle que soit la consigne ($\alpha = 0,024$).

Les deux consignes ne semblent pas produire d'évaluations significativement différentes d'un même produit, contrairement au cas de l'évaluation monadique, mais les produits sont mieux discriminés pour une même consigne. Puisqu'il n'y a que 5 essais (2 pour la consigne sportivité et 3 pour la consigne niveau de gamme) qui ont abouti sur le choix de 207-standard, il n'est pas possible de comparer le comportement visuel des individus ayant choisi la 207-sport des individus ayant choisi la 207-standard, car cet effectif n'est pas suffisant pour effectuer des analyses statistiques selon le choix effectué.

Une dernière observation est de vérifier si les notes sont différentes selon la position du produit. On effectue un test de Student non appareillé sur les deux distributions de notes obtenues par 207-sport positionné à droite et à gauche. Les deux distributions ne sont pas significativement différentes. Enfin on observe si les notes de l'habitacle de droite sont différentes des notes de l'habitacle de gauche par un test de Student non appareillé et il n'y a pas de différences.

✓ *Produit C6*

On observe tout d'abord par consigne si l'un des habitacles est significativement plus choisi que l'autre (Tableau 8).

produit / consigne	C6-noire	C6-blanche
La plus raffinée	16% (3/19)	84 % (16/19)
La plus haut de gamme	29% (6/21)	71 % (15/21)

Tableau 8 Modalité visuelle la plus fréquemment choisie

Pour les deux consignes, la probabilité de la loi binomiale montre que la C6-blanche est significativement plus choisie que l'autre ($p < 0,05$ [$p = 0$ & $p = 0,013$]).

Le Tableau 9 présente les moyennes des notes.

produit / consigne	C6-noire	C6-blanche
La plus raffinée	35	65
La plus haut de gamme	47	52

Tableau 9 Moyenne des notes sur 100 pour chaque habitacle

Un test de Student non appareillé montre que les notes de C6-noire sous les deux consignes ne sont pas différentes ($\alpha > 0,05$). Le résultat est le même pour C6-blanche. Un second test de Student appareillé cette fois montre que les notes de C6-noir et de C6-blanche sont significativement différentes pour la consigne de raffinement ($\alpha = 0,036$). Pour la consigne niveau de gamme, les notes ne discriminent pas les produits ($\alpha > 0,05$). Ici aussi on compare les notes obtenues par l'habitacle le plus choisi selon sa position droite ou gauche dans l'image et le test de Student ne dégage aucune différence.

Il y aurait une différence entre la décision et l'évaluation. En effet la C6-blanche est choisie à la fois comme la plus raffinée et la plus haut de gamme par rapport à la C6-noire. Cependant, au niveau des notes, si le caractère raffiné de C6-blanche est confirmé car les notes sont meilleures que celles de C6-noire, la C6-blanche elle n'obtient pas de notes significativement différentes de celles de C6-noire pour la consigne niveau de gamme. Il y a donc décorrélation entre le choix fortement orienté et les notes très peu discriminantes. On pourrait être en présence d'une situation où la décision est aisée mais l'évaluation complexe.

7.2.2 Analyse descriptive des fonctions de distribution spatiales des fixations oculaires

Cette première partie d'étude consiste à établir les caractéristiques des fonctions de distribution spatiales des fixations oculaires pour chaque condition expérimentale, afin d'obtenir des premières informations sur les tendances comportementales afférentes à l'évaluation visuelle par paire.

7.2.2.1 Observation de la distribution spatiale des points de fixation

L'établissement de la densité de Parzen permet d'observer les tendances du comportement du regard communes aux participants d'une même condition expérimentale.

✓ *Produit 207*

La Figure 40 présente les densités de Parzen pour chaque paire et consigne du produit 207.

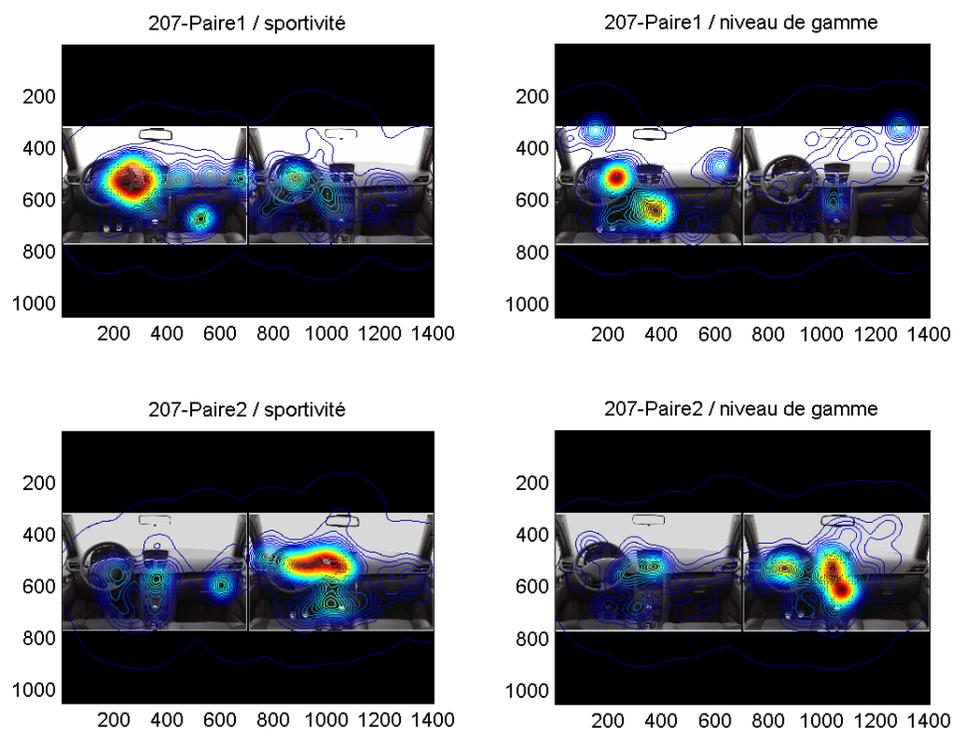


Figure 40 Densités de Parzen pour le produit 207

C'est ainsi que l'on observe pour le cas « 207-paire1 / sportivité » que la densité des fixations est plus importante sur l'habitacle de gauche. Il s'agit de l'habitacle 207-sport. L'observation de l'image semble se faire principalement sur la zone volant/haut de console centrale, et le mode le plus dense (le plus rouge, aussi nommé mode principal) n'est pas centré sur un objet précis mais à mi-chemin entre la console centrale et le volant. On observe de plus que le volant de l'habitacle de droite reçoit

des fixations, et plus spécifiquement le logo. On constate que l'exploration de l'image est plutôt sur des zones locales que dispersée sur la totalité de la surface. Enfin, on note que l'habitacle choisi est aussi celui qui est le plus observée (207-sport).

Les densités de Parzen modélisant les distributions des fixations oculaires pour chacune des 4 conditions expérimentales de 207 ont pour caractéristique commune d'être asymétriques sur la scène, avec l'existence de modes latéraux et non pas uniquement un mode central. On constate de plus que le mode principal est centré entre le volant, le compteur et la console centrale, sauf pour le cas « 207-paire2 / consigne niveau de gamme » où le mode principal est sur la console centrale. Le levier de vitesse de la 207-sport est toujours inclut dans un mode, ainsi que les logos. Le produit 207-sport est le plus fixé quelle que soit la consigne, et il s'agit de l'habitacle significativement choisi par l'ensemble des participants.

✓ *Produit C6*

Les graphiques suivants présentent les cartes de densité des fixations de Parzen pour les paires de C6 pour les deux consignes.

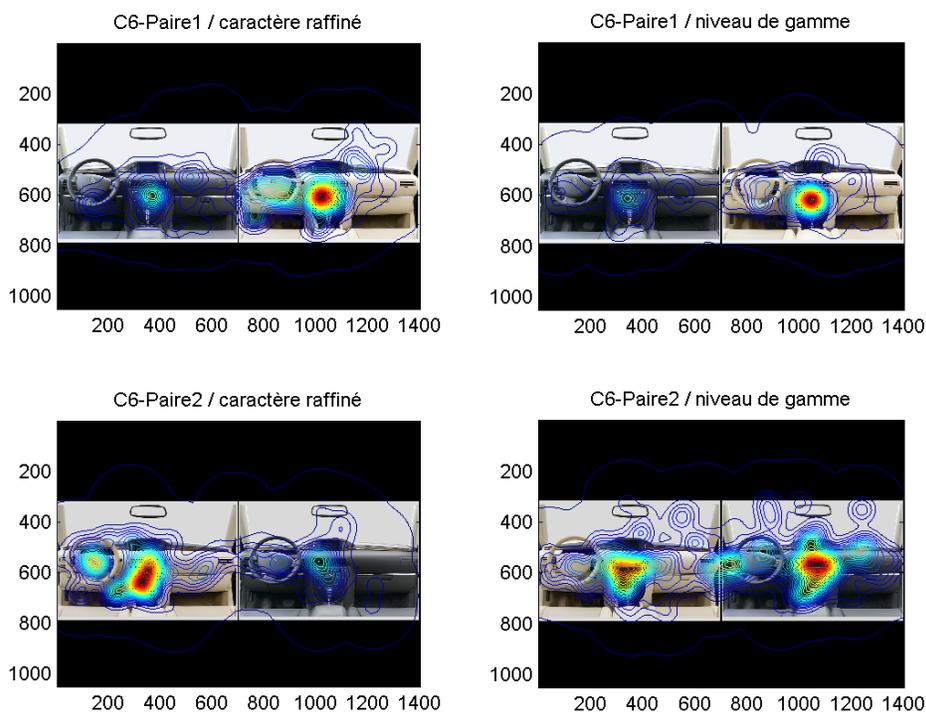


Figure 41 Densités de Parzen pour le produit C6

Nous détaillons le cas « C6-paire1 consigne caractère raffiné », puis nous présenterons une synthèse des observations. On constate qu'il y a deux modes symétriques centrés sur le centre de chaque habitacle au niveau de la console centrale. Les volants sont peu traités et les courbes de niveaux sont très étendues.

De manière générale on note sur la Figure 41 que la modalité blanche du produit C6 est la plus observée quelle que soit la condition expérimentale, et de manière très centrale, à l'exception du cas « C6-paire2/niveau de gamme ». Il existe un mode secondaire (moins dense que le mode principal)

au centre de la modalité visuelle C6-noire pour tous les cas, sauf encore une fois à l'exception du cas « C6-paire2/niveau de gamme », où le mode secondaire est sur la modalité blanche et le mode principal sur la modalité noire. Les courbes de niveaux des densités, plutôt que d'être denses comme pour le produit 207, sont plutôt étalées et suivent assez précisément les contours de la planche de bord.

En synthèse, le traitement visuel dont bénéficie le produit 207 est différent de celui de C6. Le premier est plutôt local, le second plutôt ambiant. Dans tous les cas (sauf un), l'un des éléments de la paire est plus observé que l'autre, et c'est la modalité la plus choisie et la mieux notée qui est la plus observée. Nous allons ensuite étudier l'évolution de l'orientation du regard au cours du temps afin d'observer s'il existe des phases temporelles de comportement du regard.

7.2.2.2 Evolution de l'orientation du regard au cours du temps

Cette étude consiste à calculer la densité spatiale des fixations oculaires par fenêtrage temporel par l'estimation des densités de Parzen, et à les comparer ensuite pour dégager les caractéristiques de l'évolution du déploiement de l'attention sur la scène visuelle. Les différents résultats sont présentés en annexe (annexe A-1) et sont commentés ici.

✓ *Produit 207*

Afin d'illustrer notre propos, la figure suivante représente pour le cas « 207-paire1 / consigne sportivité », l'évolution de la densité des fixations pour 2, 4, 6, 8, 10 et 12 secondes d'exposition.

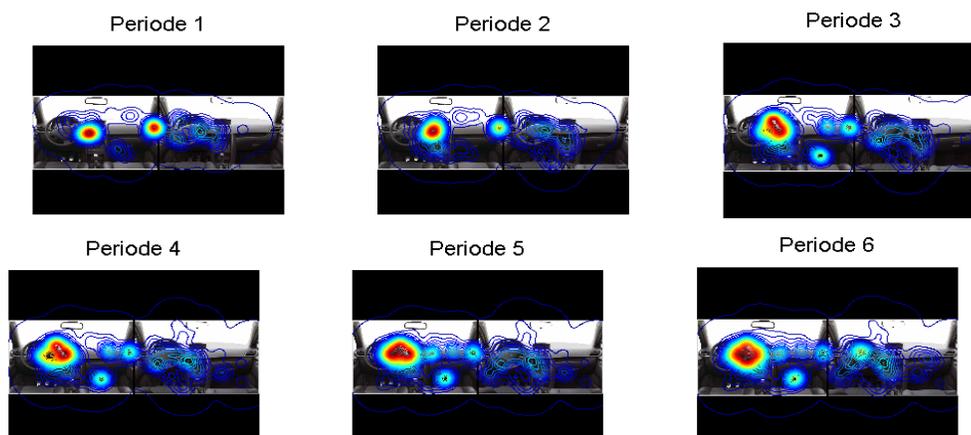


Figure 42 Observation temporelle des densités de Parzen pour la condition expérimentale « 207-paire 1, consigne sportivité »

On observe sur ce cas (Figure 42) que dans les deux premières secondes, le regard est positionné sur la zone centrale de l'image et sur le centre de la modalité visuelle de gauche (celle qui sera choisie ensuite). Puis le mode central disparaît, et l'observation de la zone compteur/volant se renforce. A la 8^{ième} seconde (période 4), le regard a traité la modalité visuelle de droite mais se stabilise sur celle de gauche, les cartes de densité n'évoluant que peu jusqu'à la fin.

Ainsi, pour le produit 207, deux types de dynamiques sont observées, et ce selon la consigne. Pour la consigne sportivité, la modalité visuelle qui sera choisie est sur-fixée dès les premières secondes. Il apparaît ensuite un mode central à la scène et un mode sur l'habitacle choisi. Puis le mode central disparaît, alors que le mode sur l'habitacle choisi se renforce ; enfin il émerge un mode secondaire sur l'habitacle rejeté au niveau du centre de la modalité visuelle et en direction du volant.

Pour la consigne niveau de gamme, un processus plus net de comparaison est visible. Il naît tout d'abord un fort mode sur la modalité visuelle qui sera choisie, puis une phase de comparaison avec des modes qui se positionnent symétriquement sur chaque modalité, et enfin un renforcement de l'observation de la modalité choisie.

Pour la consigne de sportivité, tout se passe comme si le fait de reconnaître très tôt un typage sport sur l'une des modalités visuelles inhibait la phase de comparaison locale au profit d'une confirmation par observation globale de la modalité pré-choisie. Tandis que pour la consigne « niveau de gamme », une première impression apparaît mais n'est pas suffisamment robuste pour inhiber une phase de comparaison locale qui servira à confirmer la première impression.

✓ *Produit C6*

La Figure 43 présente les cartes de densité de Parzen par pas de 2 secondes pour le cas C6-paire1, consigne « la plus raffinée ».

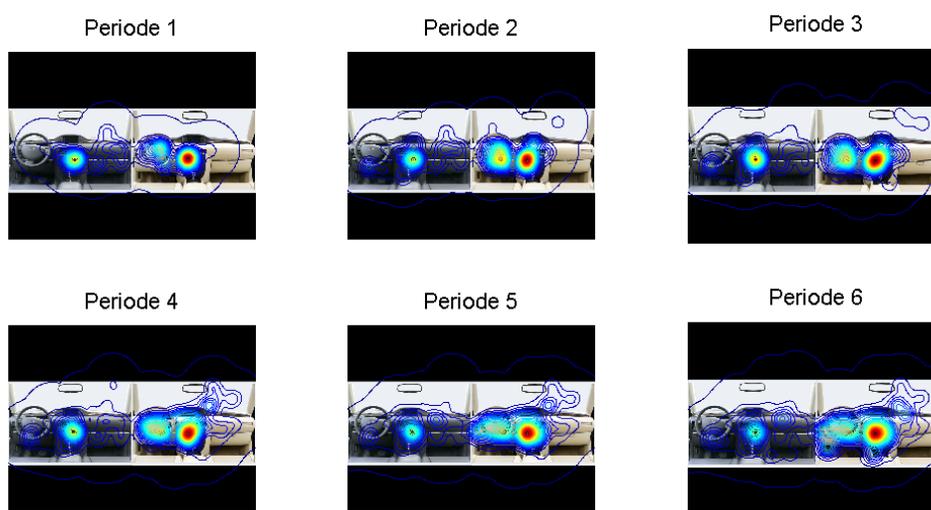


Figure 43 observations temporelle des densités de Parzen pour la condition expérimentale « C6-paire 1, consigne 'la plus raffinée' »

Pour les quatre conditions (les deux paires de C6 et les deux consignes, la plus raffinée et la plus haut de gamme), dans les deux premières secondes, il apparaît un mode principal au centre de la modalité visuelle (habitable) qui sera choisie ainsi qu'un mode secondaire au centre de la modalité qui sera rejetée. Puis, le mode principal se renforce. Autour des deux modes centraux à chaque modalité se déploie progressivement l'exploration visuelle des deux habitacles. Pour la paire 1, le regard semble attiré par la modalité blanche qui est la plus proche du point de fixation d'origine du regard (pointeur de pré exposition). Les yeux n'effectuent pas d'exploration détaillée de l'autre modalité car il y a cohérence entre la modalité la plus proche du regard (celle de droite) et un pré-choix (pré-choix de l'habitable de droite); il y aurait donc inhibition de la comparaison. De fait, la modalité blanche est préférentiellement choisie. Pour la paire 2, le regard est initialement plus proche de la modalité noire (car elle est positionnée à droite), donc s'y oriente favorablement pour des raisons spatiales. Cependant la modalité blanche est attractive ; il y a alors activation d'un processus de comparaison entre la modalité visuelle qui a été observé le plus aisément (C6-noire par proximité spatiale), et la modalité dont les attributs perceptifs pré-orientent le choix (C6-blanc). Cependant, parce que les différences entre notes ne reflètent pas fidèlement le choix, on peut considérer que les critères de jugements sont très variables et donc que l'exploration visuelle ne peut pas mettre en évidence la recherche d'informations visuelles locales spécifiques.

✓ *Synthèse des observations des fonctions de distribution de densité de Parzen*

Le paradigme de comparaison par paire pour des habitacles de véhicules met en évidence différentes phases temporelles de traitement visuel : pré-décision, comparaison, confirmation. On observe qu'il y a une sur-observation de la modalité qui sera choisie et que selon la modalité visuelle ou la consigne, la prise d'information sera plutôt ambiante ou fovéale. Ces résultats ne prédisent pas la pertinence du traitement par zones d'intérêts mais démontrent d'ores et déjà que de nombreuses informations sur les processus visuels peuvent être extraites de ce type de paradigme.

L'estimation des densités par Mélange de Gaussienne va permettre d'aller plus loin dans la compréhension des processus de prise d'information visuelle, en caractérisant les modèles de distribution des fixations.

7.2.2.3 Estimation des fonctions de densité par Mélange de Gaussiennes

L'établissement des fonctions de distribution spatiale des fixations oculaires par la méthode du Mélange de Gaussienne demande à être initialisée avec une liste de centres spatiaux. Cette liste de centres est fournie par un calcul de barycentres par méthode de Meanshift et est ensuite corrigée manuellement si nécessaire. Cette démarche a été présentée précédemment dans le chapitre « méthode » ainsi que lors de l'expérimentation 1. Cette méthode de modélisation par mélange de gaussiennes est appliquée afin de dégager l'existence ou l'absence d'un mode uniforme captant les fixations liées à la nature aléatoire de la distribution, d'établir des indicateurs décrivant le nombre et les positions des modes, ainsi que leur orientation et leur intensité. L'ensemble des figures sont placés en annexe (annexe A-2).

✓ *Modélisation pour le cas » 207-paire 1 / consigne sportivité »*

La Figure 44 permet d'effectuer le choix du meilleur modèle.

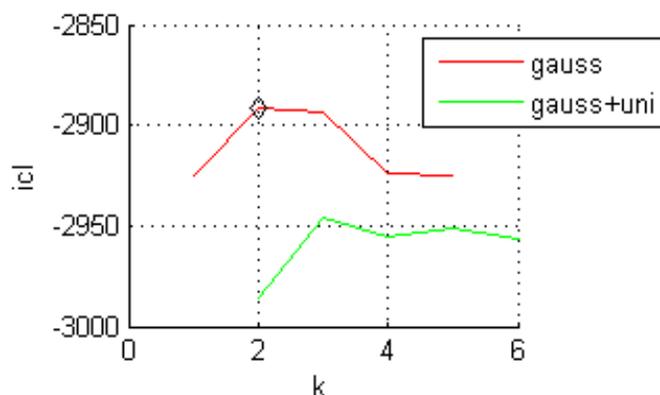


Figure 44 Evolution du critère d'information en fonction du nombre de modes

On observe que le meilleur modèle comporte deux modes, sans mode uniforme.

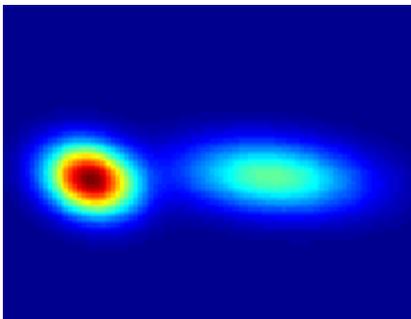


Figure 45 Densité des fixations

On constate sur la Figure 45 que le mode de densité de gauche est plus intense que le mode de densité de droite.

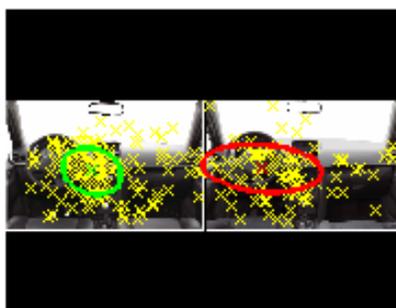


Figure 46 Contours des modes à un écart type

Cette représentation (Figure 46) permet de dégager que le mode de gauche, situé au centre de la modalité visuelle contenant l'habitacle qui sera choisi, a une plus faible variance que celui qui est positionné sur la modalité visuelle de droite.

✓ *synthèse*

On constate que les distributions de fixations de toutes les conditions expérimentales sont modélisées par deux modes, sans modes uniformes, et que ces modes sont centrés sur chaque modalité visuelle. L'évaluation par paire se ferait donc depuis un point de vue central à chacune des deux modalités visuelles ; cela explique la contribution nulle d'un mode uniforme. De manière cohérente, chaque modèle est systématiquement constitué de deux modes. Cependant, le mode sur la modalité visuelle choisie est le plus dense, indiquant que l'habitacle choisi est plus observé, et le mode de la modalité visuelle droite est plus étendu, donc traité moins localement, sauf pour le cas « C6-Paire2 consigne niveau de gamme ».

Cette analyse des fonctions de densité des fixations montre que l'évaluation par paire génère un traitement visuel très différent de l'évaluation monadique. Il existe une relation entre le temps passé à observer chaque modalité visuelle et le choix, ainsi qu'un fort processus d'observation depuis une position centrale à chaque modalité visuelle. Enfin, on distingue différentes phases temporelles de stratégies visuelles. Afin d'aller plus loin dans l'analyse de ces données, nous allons étudier en détail les similarités et différences entre les modèles de densité.

7.2.3 Analyse comparative des distributions de fixations

✓ *Similarités spatiales des distributions de fixations*

Après avoir observé de manière descriptive les caractéristiques des distributions de fixation par condition expérimentale, on va chercher ici à les comparer afin de mettre en évidence si oui ou non les facteurs expérimentaux (consigne, image) expliquent ces distributions, donc si l'orientation du regard est homogène intra-condition et guidée par la consigne ou l'image. Pour cela, on va employer la métrique de Kullback-Leibler, définie précédemment dans le chapitre méthode, sur les cartes de densités de Parzen. Toutes les fixations de tous les essais d'une même condition sont combinées pour établir la densité de Parzen. On calcule ensuite la distance de Kullback-Leibler entre toutes ces cartes de densité, pour établir une matrice de dissimilarité, sur laquelle on effectue une Classification Ascendante Hiérarchique pour regrouper les modèles qui sont les plus proches. Puisque les modélisations par mélanges de gaussiennes ont montré pour toutes les conditions expérimentales une configuration optimale du modèle avec deux modes gaussiens (un sur chaque modalité visuelle), alors une classification des modèles de densité selon leur distance de Kullback-Leibler devrait produire un groupe avec les cas dont l'habitacle choisi est à gauche de l'image et un groupe de modèles où l'habitacle choisi est à droite.

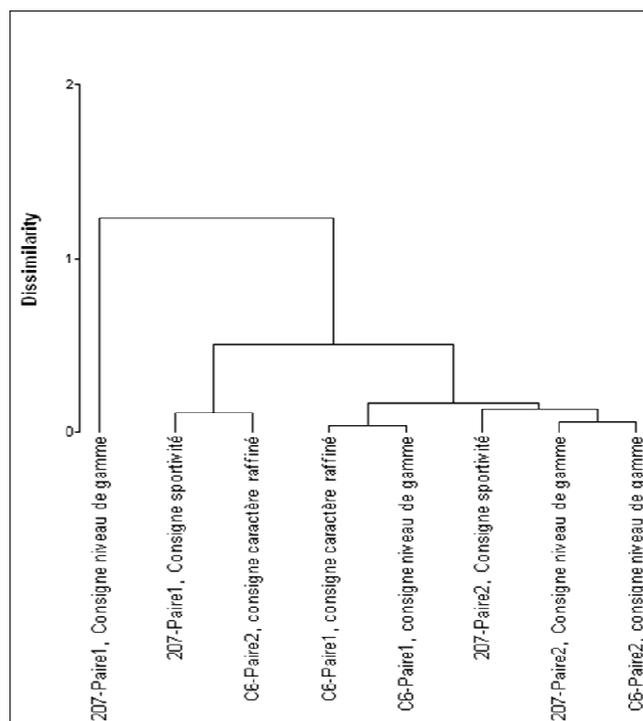


Figure 47 Résultats de la Classification Ascendante Hiérarchique de la distance de Kullback-Leibler entre modèles de densité de Parzen

La classification illustrée Figure 47 montre que les distances entre modèles ne sont pas spécifiquement plus faibles pour une même paire sous deux consignes, que pour deux paires différentes sous une même consigne. On n'observe pas non plus de groupes selon la position de la modalité visuelle choisie. Ni les conditions expérimentales ni les résultats évaluatifs ne semblent justifier la composition des groupes de fonctions de densité. Cela peut s'expliquer par l'hypothèse que pour ce type de paradigme, quelles que soient les modalités visuelles ou la consigne, l'observation du regard se fait principalement depuis un point de vue central à chaque habitacle (processus de maximisation de l'information par exemple). La différence entre modèles serait alors due à

l'hétérogénéité des essais intra-modèles. Ce résultat devrait se retrouver dans l'étude des zones d'intérêt ou les zones traitées ne devraient pas être spécifiquement dépendantes des conditions expérimentales. On va ainsi chercher à tester l'homogénéité des essais. L'une des méthodes possible est d'établir la densité moyenne par essai sur les différents modèles de densité.

✓ *Homogénéité intra-modèles*

L'objectif de ce traitement est de tester si le modèle de densité reflète bien les essais qui le compose, donc l'homogénéité de traitement spatial par condition. En effet, si les participants, pour une même condition, observent des zones très différentes, alors le modèle de densité établi comportera peu de modes de forte densité, mais plutôt des modes à faible variance ; il ne sera alors pas représentatif d'un processus d'orientation du regard homogène pour une condition expérimentale. Si au contraire les participants observent de nombreuses zones communément, alors il apparaîtra des modes de forte de densités bien localisés. Cela ne signifiera pas que c'est la consigne ou l'image qui aura produit ce positionnement homogène, mais qu'il y a un processus de prise d'information commun. A l'inverse, si l'on ne dégage pas d'explication par les conditions expérimentales aux distributions des fixations, alors on peut avancer que l'homogénéité des participants est faible, donc que l'exploration de la scène est peu guidée par les conditions expérimentales (image, consigne).

Une matrice de confusion est ainsi établie selon le nombre de fois qu'un essai est plus vraisemblable d'appartenir à une autre condition expérimentale que celle dont il est issu, sur la base de la densité moyenne des lieux de fixations sur les différents modèles de densité. Si l'homogénéité des essais est bonne par conditions expérimentales, il ne doit pas y avoir de confusion dans la vraisemblance d'appartenance. La matrice de confusion est représentée par le Tableau 10.

	A	B	C	D	E	F	G	H
A	91	0	0	0	0	9	0	0
B	0	100	0	0	0	0	0	0
C	11	0	56	22	0	11	0	0
D	11	0	11	56	11	11	0	0
E	13	0	0	0	75	13	0	0
F	0	0	0	0	0	100	0	0
G	0	0	0	0	0	0	100	0
H	0	0	0	0	0	0	25	75

A	207-Paire1, Consigne sportivité
B	207-Paire1, Consigne niveau de gamme
C	207-Paire2, Consigne sportivité
D	207-Paire2, Consigne niveau de gamme
E	C6-Paire1, consigne caractère raffiné
F	C6-Paire1, consigne niveau de gamme
G	C6-Paire2, consigne caractère raffiné
H	C6-Paire2, consigne niveau de gamme

Colonnes : modèle d'origine de l'essai ; lignes : modèle d'attribution de l'essai

Tableau 10 Matrice de confusion : pourcentage de bonnes attributions.

Pour le produit 207, il semble que la confusion soit plus forte pour la paire 2. Dans la paire 2, l'habitable 207-sport est positionné à droite, et c'est celui-ci qui est majoritairement choisi quelle que soit la consigne. Lorsque le biais d'observation centrale se combine avec l'attractivité de la modalité visuelle choisie (poste de conduite de l'habitable de droite proche du centre de la scène), un fort centre d'attraction spatial apparaît sur la modalité visuelle de droite, et la capacité d'une zone à attirer

le regard diminue lorsque sa distance à ce centre augmente; l'exploration devient donc déséquilibrée car si tous les participants observent bien ce centre d'attractivité, rien n'oriente leur regard lorsqu'ils observent la seconde modalité visuelle. A l'inverse, pour la paire 1, lorsque le biais d'observation central vient contrebalancer l'attractivité de l'habitacle qui sera choisi, deux centres de gravité se forment et stabilisent le regard sur chacun des deux habitacles; l'exploration devient alors plus homogène car le regard se pose sur l'un des deux centres. Cela se manifeste lorsque l'exploration est plutôt de type focal, car il existe plusieurs centres d'intérêt potentiels sur lesquels le regard peut se poser.

Ce phénomène ne s'observe pas pour les produits C6, qui ne possèdent que peu de centres locaux d'intérêts. Dans cette situation, la confusion n'est liée ni à la composition de la paire, ni à la consigne. Nous ne proposons pas d'explication à ces mauvaises affectations pour C6, hormis la prégnance d'une composante aléatoire de l'exploration visuelle.

7.2.3.1 Conclusion sur l'analyse des fonctions de distribution des densités

Les résultats présentés permettent de dégager les principales tendances du déploiement de l'attention sur la scène. A ce stade de l'étude les conditions expérimentales (consigne ou modalité visuelle), ne permettent pas d'expliquer le comportement du regard. Tandis que ces observations sur l'attention visuelle ont été effectuées à une échelle macroscopique, la section suivante propose une analyse plus précise en cherchant à discriminer chacune des deux modalités visuelles par le traitement oculaire dont elles bénéficient.

7.2.4 **Analyse des processus temporels de l'attention visuelle sur la scène**

L'observation de l'évolution temporelle des fonctions de densité de Parzen permet de poser l'hypothèse de l'existence de stratégies visuelles évoluant en fonction du temps. En effet, dans certaines situations, le regard semble refléter une pré-décision, puis une exploration comparative, et enfin une stabilisation orientée. Il est ainsi intéressant de tester si des descripteurs du comportement oculomoteur confirment l'existence de ces différentes phases, permettant de mieux comprendre les processus de cognition visuelles activés par l'évaluation d'un produit manufacturé.

L'observation du taux d'impacts²⁰ sur l'élément choisi en fonction du temps va confirmer ou non si le regard s'oriente effectivement favorablement vers la modalité visuelle qui sera choisie. L'observation de la durée des fixations en fonction du temps permet de vérifier s'il existe des phases dans lesquelles le regard se stabilise. Enfin, une Analyse par Composante Principale des variables « 'timecode'²¹ de début de fixation », « amplitude de saccade », et « durée de fixation » sur les individus « fixations » permet d'observer s'il existe des groupes de fixations discriminés par ces dimensions.

7.2.4.1 Taux d'impacts sur la modalité visuelle choisie au cours du temps

A chaque instant de l'exposition, pour tous les essais d'une condition expérimentale, la position du regard est relevée en termes d'orientation. On établit ainsi la fonction $f(t)=1$ si le regard est posé sur l'élément de la paire qui sera choisi, $f(t)=0$ sinon. Le taux d'impacts sur l'élément choisi est ainsi établi pour chaque instant de l'exposition sur l'ensemble des essais. S'il existe dans l'orientation du regard

²⁰ Les impacts sont les mesures de la position du regard échantillonnées à la fréquence du système de capture, données de bases au calcul des fixations et saccades.

²¹ Le timecode est le relevé temporel en seconde du déroulement de l'exposition, valant 0 au début de l'exposition et 12s à la fin

une tendance commune à tous les essais, ce taux être élevé, indiquant une orientation homogène significative. Dans l'hypothèse où il n'y a pas de cohérences inter-essais, ce taux va varier similairement à une oscillation aléatoire. Les figures résultantes sont présentées en annexe B-1.

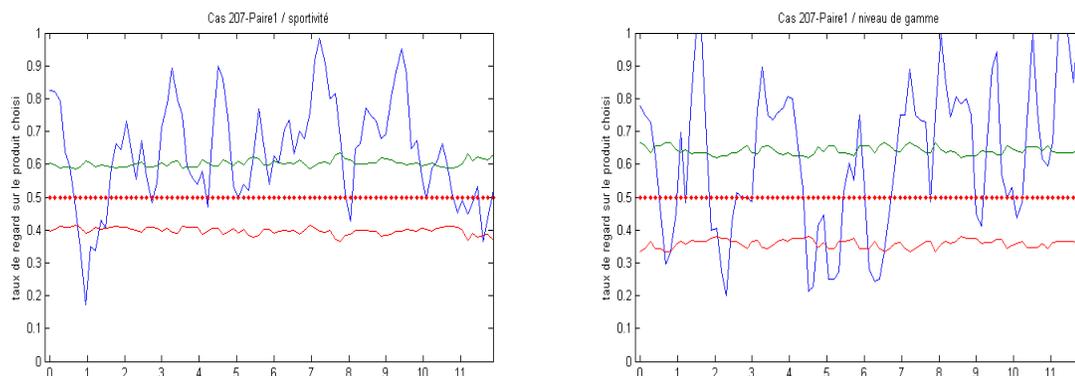


Figure 48 Taux de points de regard sur l'habitable choisi. En bleu, la courbe principale oscillant entre 0 (pas de points de regard sur l'habitable choisi) à 1 (tous les points de regard du moment positionnés sur l'habitable choisi). En vert, le taux supérieur de la loi binomiale au-delà duquel le taux mesuré est significativement au dessus du hasard et en rouge le taux inférieur en deçà duquel le taux mesuré est significativement en dessous du hasard.

Sur les deux cas exposés Figure 48, « 207-paire 1 / consigne sportivité » et « 207-paire 1 / consigne niveau de gamme », nous dégageons les observations suivantes. Pour le « cas 207-paire 1 / consigne sportivité », l'habitable choisi est observé dans la première demi seconde, puis l'habitable rejeté, et le regard s'oriente ensuite très nettement (presque tout le temps au dessus du seuil de significativité) vers cette modalité visuelle « sport ». On remarque qu'à la fin le taux n'est pas significatif d'une orientation. Pour le cas « 207-paire 1 / consigne niveau de gamme », on constate aussi que l'habitable choisi est significativement sur-observé dans la première demi-seconde. Puis, le regard va osciller entre les deux modalités visuelles sans privilégier l'une ou l'autre, cependant, à partir de la 7^{ème} seconde, le regard va s'orienter de plus en plus favorablement vers l'habitable qui sera choisi. Tout se passe comme si dans le cas de l'évaluation de la sportivité le choix était pré effectué dans la première seconde, orientant ensuite l'investigation détaillée de la modalité visuelle pré-choisie, confortant ce choix, et produisant finalement un effet aléatoire en fin d'essai par fin de tâche avant la fin de l'exposition. Pour la consigne niveau de gamme, un pré choix est aussi effectué mais étant moins certain que dans le cas précédant. Une investigation comparative va alors avoir lieu, renforçant progressivement le pré-choix effectué et orientant de plus en plus le regard vers cette modalité visuelle pour une investigation locale.

En synthèse le taux de regard démontre l'existence de phases temporelles. Ce taux montre une oscillation du regard d'une modalité visuelle à l'autre, s'orientant de plus en plus significativement en direction de l'élément qui sera choisi au final. Pour les consignes où les deux modalités sont fortement discriminées (sportivité et raffinement), le regard oscille moins que dans les cas où les produits sont moins discriminés (niveau de gamme). En terme temporel, l'orientation du regard dans la première seconde apparaît pour 3 des cas de l'étude du produit 207 (tous sauf 207-paire2, consigne sportivité) s'orienter significativement vers l'élément qui sera choisi puis s'oriente en direction de l'autre. Cela pourrait être la démonstration d'un processus de pré-décision mais aussi d'attractivité par saillance visuelle du fait d'un choix systématiquement en faveur de 207-sport. On pourrait expliquer le cas isolé pour le produit 207 (207-paire2, consigne sportivité) par le fait que pour cette condition expérimentale la modalité la plus sportive soit à droite, donc juste à coté du point d'origine du regard, et que la consigne (sportivité) oriente rapidement la choix vers cette modalité. Dans ce cas, la

dépense d'énergie d'un déplacement manifeste du regard en direction de cet habitacle n'est peut être pas jugée utile, puisque l'information peut être extraite aisément depuis le point central.

Pour C6, le regard ne s'oriente pas significativement dès le début de l'exposition vers l'habitacle qui sera choisi, sauf pour le cas C6-Paire1 consigne niveau de gamme. Pour le produit C6, le cas isolé est celui où l'habitacle choisi est à droite et où le choix n'est pas clairement en faveur d'un des deux produits (C6-Paire1 consigne niveau de gamme). Le regard s'oriente significativement vers la modalité noire qui sera ensuite rejetée. Il pourrait s'agir d'un processus de comparaison rapide pour essayer de trouver des différences visuelles significatives pour la consigne. Enfin, la dernière et non négligeable des explications sur la variabilité des résultats expérimentaux pourraient provenir d'un effet de groupe de nature aléatoire, c'est-à-dire d'un manque de robustesse des données, limitant la significativité des résultats. L'augmentation du nombre d'essais sur la condition expérimentale permettrait d'améliorer la significativité des tendances.

Cette section ayant permis de mettre en évidence l'existence de phases temporelles, nous allons analyser de manière mono et multi variée les comportements oculomoteurs temporels.

7.2.4.2 Analyse des caractéristiques temporelles de la prise d'information

Décrire le comportement oculomoteur en fonction de la modalité visuelle observé peut permettre de dégager des tendances communes à tous les essais. Les variables observées sont les suivantes :

- Le taux de fixations sur l'habitacle choisi en fonction du temps est établi car permet d'observer le lien entre le comportement visuel et la décision
- La durée moyenne des fixations et l'amplitude des saccades en fonction du temps sont aussi extraites car peuvent indiquer l'existence de phases temporelles d'activité ou de stabilité du regard
- Le taux de saccades intra-modalité visuelle est établi car permet de distinguer les phases d'exploration d'une modalité visuelle et les phases de comparaison.

Ces variables sont établies par rang de fixation et non du temps, car chaque fixation n'ayant pas la même durée, il ne serait pas possible de comparer différents essais. Puisque tous les essais ne comportent pas le même nombre de fixations, l'analyse des rangs s'arrête lorsque 60% des essais sont achevés (c'est-à-dire en moyenne lorsqu'il ne reste plus que 3 essais non achevés).

✓ *Résultats Produit 207*

▪ *tendances des variables*

Les résultats sont placés en annexes B-2.a à B-2.d. On nomme figure chacune des deux modalités visuelles qui composent la paire. Pour chaque condition expérimentale on observe les tendances de chacune des 4 variables définies précédemment.

- cas 207-paire1, consigne sportivité

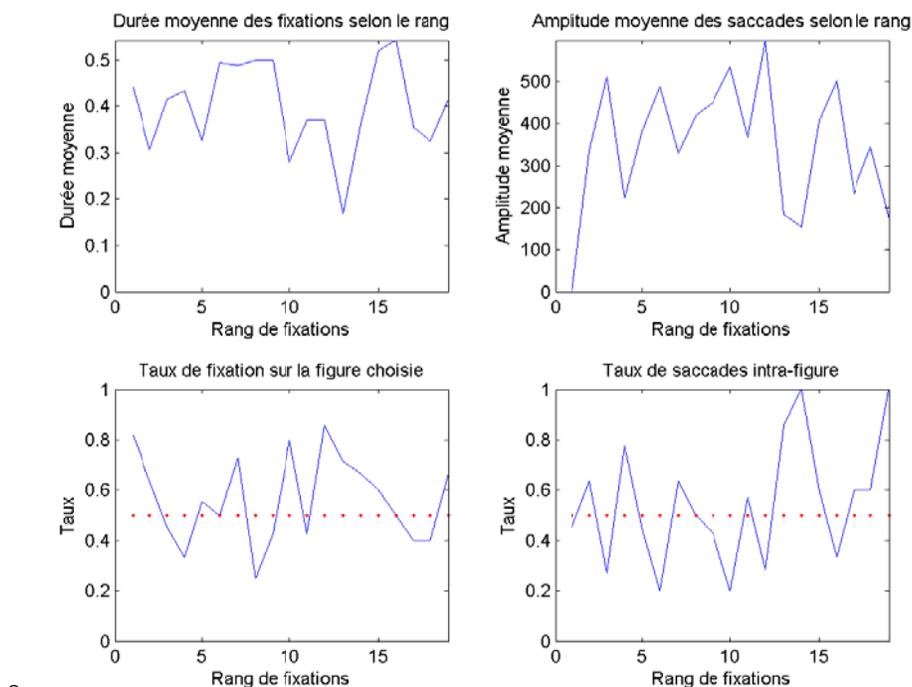


Planche 3 Variables décrivant les caractéristiques des fixations en fonction du rang

Il n'apparaît pas, sur la planche 3, de tendance dans les durées de fixations en fonction du rang. Le taux de fixations sur la figure choisie indique que le regard est significativement orienté vers cette dernière au début de l'exploration, puis on observe une oscillation avec une orientation en sa faveur. Cela est confirmé par le taux de fixations de type intra-figure, au début plutôt faible, et augmentant avec le temps.

- cas 207-paire1, consigne niveau de gamme : le taux de fixation sur la figure choisie et le taux de saccades de type intra montrent une forte oscillation du regard entre les deux figures, avec en fin de séquence une augmentation du taux de saccades intra, logiquement combiné avec une décroissance de l'amplitude des saccades.
- cas 207-paire2, consigne sportivité : le seul indicateur montrant une tendance est celui du taux d'observation de la figure choisie, le regard se dirigeant de plus en plus vers celle-ci au cours du temps.
- cas 207-paire2, consigne niveau de gamme : on observe que les durées de fixations diminuent avec le rang, alors que la quantité de saccade intra aurait tendance à augmenter.

▪ *Corrélations entre variables*

On étudie ensuite les corrélations entre les 4 variables définies. Au niveau des corrélations significatives entre variables, on observe communément aux quatre conditions expérimentales une anti corrélation amplitude moyenne / taux de fixation intra figure : cela signifie que lorsque le taux de fixation intra figure est fort, l'amplitude des saccades est faible, ce qui est cohérent d'un point de vue spatial. Il existe de plus la corrélation suivante commune à 207-paire1 consigne niveau de gamme, et 207-paire2 consigne sportivité : corrélation entre fixation intra figure et rang : plus le temps d'exposition augmente, moins il y a comparaison des figures. Ces deux cas n'ayant en commun ni la consigne ni l'image, il n'est pas possible d'en déduire la présence d'un processus spécifique.

Les corrélations propres à chaque condition sont discutables par un potentiel manque de robustesse des données.

Pour tous les cas, les variables taux de fixations intra-modalité et durée moyenne des fixations sont indépendantes. Cela signifie que l'exploration intra-modalité n'implique pas des fixations de durées significativement plus courtes ou longues que l'exploration inter-figure.

Les variables rang et durée moyenne des fixations sont indépendantes pour tous les cas sauf 207-paire2 consigne niveau de gamme. Cela signifie que les durées de fixations ne sont pas plus courtes ou plus longues selon le temps d'exposition, ne permettant donc pas d'appuyer l'hypothèse d'une stabilisation du regard avec le temps.

Une Analyse par Composante Principale ne permet pas de dégager de groupes de rang en fonction de ces variables, qui auraient pu être associés à des phases temporelles cognitives.

✓ *Résultats Produit C6*

Les données sont placées en annexe B-2.e à B-2.h.

▪ *Tendances des variables*

- Cas C6-paire1 (modalité choisie à droite), consigne raffinement : il semble que la durée moyenne des fixations diminue avec le temps ainsi que l'amplitude des fixations, alors que le taux de saccade intra modalité et le taux de fixation de la figure choisie augmente.
- Cas C6-paire1 (figure choisie à droite), consigne niveau de gamme : il ne semble pas y avoir de tendances temporelles spécifiques.
- Cas C6-paire2 (figure choisie à gauche), consigne raffinement : il semble y avoir une forte variabilité selon le rang, ce qui pourrait être dû à une variabilité intra essai mais aussi à une variabilité inter essais. Il ne semble pas qu'il existe de zones stabilisant le regard.
- Cas C6-paire2 (figure choisie à gauche), consigne niveau de gamme : pour ce cas non plus il ne semble pas qu'il existe un fort effet d'homogénéité selon le rang. De grandes oscillations existent.

▪ *Corrélations entre variables*

Uniquement pour la paire 2 quelle que soit la consigne, on note une corrélation rang / moyenne des durées de fixation : plus le rang de fixation augmente, plus les durées des fixations sont longues. La corrélation 'taux de fixation intra' et 'amplitude de saccade' n'apparaît que pour les cas « C6-paire 1, consigne niveau de gamme » et « C6-paire 2, consigne sportivité ». Ce qui signifie que pour les deux autres cas, l'amplitude est indépendante du fait de passer d'une modalité à l'autre ou de ne regarder qu'une modalité. Les amplitudes intra figures seraient donc aussi grandes que les amplitudes inter-figures.

Une Analyse par Composante Principale ne permet pas de dégager de groupes de rang selon ces variables

7.2.4.3 Synthèse

Ces traitements ont pour but de dégager si le comportement du regard se décompose en phases temporelles (par exemple l'exploration intra modalité a lieu en fin de séquence, ou encore le taux de fixation de la figure choisie augmente avec le temps). L'analyse de ces variables pour le produit 207 ne met pas en évidence de stratégies selon les conditions expérimentales. Il n'est donc pas possible

d'associer les variables « durée des fixations, amplitude des saccades, taux de fixation du produit choisi et taux de saccades intra-modalités » à un comportement visuel spécifique.

L'analyse des résultats pour le produit C6 ne montre pas d'effets de la consigne mais plutôt de la latéralisation des modalités visuelles. Dans certains cas (dont les conditions expérimentales varient), les amplitudes de saccades sont aussi importantes inter figures que intra figures. Cela indique une exploration ambiante et non focalisée de chaque modalité visuelle. Plus le temps d'exposition augmente, plus les durées des fixations sont longues, comme si le regard se stabilisait progressivement. Enfin, il n'apparaît pas que la modalité visuelle qui sera choisie soit de plus en plus fréquemment observée.

Cette étude n'a pas permis de mettre en évidence des stratégies comportementales oculaires temporelles. Des tendances semblent pourtant se dégager sans qu'il soit possible de les relier à des processus identifiés. La similarité des produits étudiés, la densité des informations visuelles et le caractère « non focal » des consignes semblent limiter l'apparition de stratégies visuelles homogènes entre participants. Cette méthode apparaît cependant comme pertinente pour vérifier des hypothèses sur l'existence de phases temporelles dans le comportement oculaire.

Dans une démarche d'étude toujours plus locale, la section suivante présente les résultats de l'analyse du comportement du regard en fonction de zones d'intérêt.

7.2.5 Analyse du traitement visuel des zones d'intérêt

7.2.5.1 Classification des zones d'intérêts par le comportement du regard

L'analyse des zones d'intérêt a pour objectif de mettre en évidence les zones visuelles de l'image qui sont traitées de manière identique par le regard, par exemple des groupes de zones observées plus longtemps que les autres ou plus tôt dans la séquence. L'objectif final est de corréliser le comportement du regard et les informations visuelles traitées pour l'élaboration du jugement. En particulier dans le cas de la comparaison par paire, on souhaite vérifier si les zones sémantiques similaires (« jumelles ») des deux modalités visuelles sont traitées de manière identique par le regard et si la modalité visuelle choisie est visuellement traitée différemment de l'autre modalité. Les effets des consignes peuvent aussi être mis en évidence en comparant les groupes de zones car cette analyse est effectuée pour chaque condition expérimentale.

Dans le cas du paradigme d'évaluation monadique, les zones d'intérêt ont été définies *a posteriori*, c'est-à-dire sur la base des densités spatiales des fixations. Dans le cas présent, nous souhaitons vérifier si des zones « jumelles » (par exemple les leviers de vitesse des habitacles de gauche et de droite dans la paire) sont traitées par le regard consécutivement ou de la même manière. De plus, les modèles de mélange de gaussiennes ne discriminent que deux groupements de fixations sur la base des densités. C'est pourquoi nous allons définir des zones d'intérêt *a priori*, symétriques sur chacune des deux modalités visuelles. Chaque habitacle est décomposé en zones géométriques, en découpant l'image en fonction des objets visuels (voir Annexe B-3).

L'orientation de l'attention vers une zone visuelle peut être due à de nombreux facteurs perceptifs et cognitifs : zone sollicitant un traitement perceptif complexe, zone attractive par sa nature physique ou par sa pertinence sémantique, zone optimale pour le traitement visuel de la scène, zone attractive par non typicité, etc. On va chercher à mieux comprendre ces effets pour chaque condition expérimentale. Ainsi, pour chaque zone contenant plus de 1% des fixations, les variables suivantes sont calculées:

- o durée cumulée de fixation de la zone (indique la capacité de la zone à attirer le regard)

- fréquence de traitement (nombre d'entrées sur une zone dans la séquence, second indicateur de l'attractivité de la zone)
- moyenne des durées des fixations de la zone (distingue les zones fixées longuement des zones juste survolées)
- rang moyen de fixation de la zone : score de Bradley-Terry d'une zone. Plus le score obtenu par une zone est élevé, plus la probabilité que cette zone soit traitée tôt dans la séquence est élevée. Plus une zone est fixée tôt, plus son effet sur l'attention est fort.
- moyenne des amplitudes des saccades précédant la fixation de la zone : mise en évidence d'une stratégie comparative

Le choix de ces variables n'est pas exhaustif, mais permet de mettre en évidence les principales caractéristiques de l'attention, spatialement et temporellement.

Ces variables, qui décrivent les caractéristiques du comportement visuel sur chaque zone d'intérêt de l'image, sont représentés par une Analyse en Composantes Principales. Sur les nouvelles dimensions fournies par l'A.C.P. est appliqué une Classification Ascendante Hiérarchique afin d'obtenir des groupes de zones. Une Analyse Factorielle Discriminante est enfin effectuée sur les variables d'origine et la variable qualitative « numéro de classe » de la zone issue de la C.A.H. Cela doit permettre de maximiser la discrimination entre groupes de zones par les variables du comportement du regard.

Pour chaque condition expérimentale, nous allons observer les caractéristiques suivantes :

- Composition de chaque classe en termes de zones visuelles (présence de zones 'jumelles' ou non par ex.)
- variables qui discriminent les classes
- présence/absence de comportement temporel homogène entre essais
- zones d'intérêts ayant reçu une attention visuelle particulière

Ces résultats sont exposés en annexe B-4 et commentés dans la section suivante.

Puisque les deux modalités visuelles sont composées des mêmes zones, il est nécessaire de les nommer distinctivement. Pour cela, plutôt que de désigner l'appartenance de la zone à la modalité visuelle gauche ou droite dans la paire, on désigne la zone selon le critère d'appartenance ou non à la modalité visuelle choisie dans l'essai. On rappelle que pour chaque condition expérimentale, l'une des deux modalités visuelles est quasiment tout le temps choisie.

✓ Cas 207-paire1, consigne sportivité

▪ Classification

On observe sur la classification des zones que plusieurs zones « jumelles », c'est-à-dire identiques sémantiquement sur le produit 1 et 2, sont classées dans le même groupe : Volant (classe 1), planche de bord, compteurs (classe 2), pédaliers (classe 4). Il n'y a que les zones 'levier de vitesse' qui sont des zones définies comme exprimant le typage par les concepteurs et qui ne soient pas classées ensemble. Enfin, on observe que tous les groupes sont composés de zones des deux produits. Il n'y a donc pas un traitement spécifique à l'un ou à l'autre. Le traitement par paire a entraîné un traitement local similaire de certaines zones des deux produits comparés. L'analyse factorielle discriminante va permettre d'expliquer les différences entre groupes de zones

▪ Discrimination

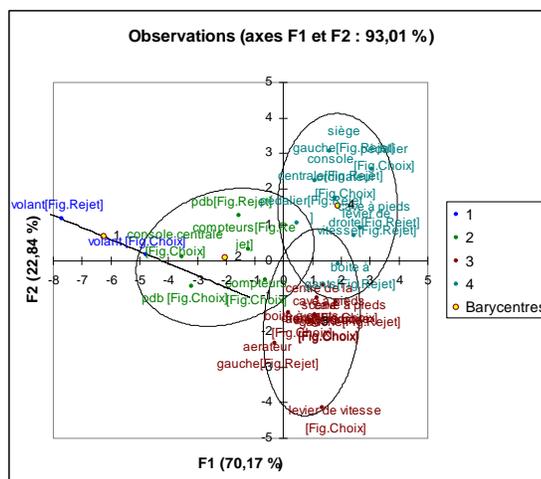


Figure 49 Résultat de l'analyse factorielle discriminante des zones d'intérêt pour le cas 207-Paire 1, consigne sportivité

La dimension 1 (70% de l'inertie) est expliquée par les variables suivantes : temps cumulé, fréquence, score de Bradley. Elle permet de discriminer la classe 'volant' et la classe contenant les compteurs et planche de bord (classes 1 et 2) des deux autres classes (3 et 4). Ces classes 1 et 2 sont spécifiquement traitées, à la fois tôt dans la séquence et fréquemment ainsi que longuement, par rapport aux classes 3 et 4.

La dimension 2 (23%) est expliquée par les variables amplitude et durée de fixation. Elle permet de discriminer les classes 3 et 4. La classe 4 (ordinateur, pédalier, console centrale) contient des zones pertinentes pour la consigne, fixées plutôt longuement en terme de moyenne de durée, et plutôt après de longues amplitudes donc significatives d'une possible comparaison avec l'autre produit. La classe 3 regroupe des zones fixées peu souvent, tardivement, avec des petites durées de fixation et de petites amplitudes. Il s'agit plutôt de zones fixées car étant proche d'une zone pertinente, sur lesquelles le regard ne revient pas.

La dimension 3 (7%) est expliquée par la durée des fixations, et permet de discriminer les classes 1 et 2, où les volants sont fixées plus longuement que les zones de la classe 2.

A ce niveau de l'analyse, on peut considérer qu'il existe 3 types de zones :

- les zones traitées tôt et souvent : volant, compteurs, planche de bord
- les zones traitées tardivement, peu de fois mais avec une orientation claire : ordinateur, pédalier, console centrale
- les zones délaissées et traitées spatialement de proche en proche

Nous allons vérifier si les résultats sont similaires lorsque la consigne est là même mais que la position des produits est inversée (deuxième paire).

✓ Cas 207-paire2, consigne sportivité

▪ Classification

Les zones console centrale, boîte à gants, et siège gauche de la figure 1 et de la figure 2 sont regroupées. On observe que les classes 1 et 4 sont plutôt composées de zones de la figure choisie tandis que la classe 2 est composée de zones de la figure rejetée, et que la classe 3 contient autant de zones de chaque figure. Il n'est pas mis ici en évidence l'existence d'un traitement sémantique focal aussi fort que sur la paire 1. Nous allons observer si les dimensions qui séparent les classes sont similaires ou non.

▪ Discrimination

La principale dimension (67% de l'inertie) qui discrimine les zones est expliquée par la variable « durée moyenne des fixations ». Elle permet de distinguer la classe 4 (ordinateurs et compteurs de la figure choisie, cave à pieds droite figure choisie), traitée avec de longues fixations, et la classe 2 (zones pédalier, planche de bord, levier de vitesse de la figure rejetée, et volant de la figure choisie) traitée avec de courtes fixations.

La seconde dimension (24% de l'inertie) est expliquée par l'ensemble des autres variables, et discrimine la classe 4 des autres classes, cette classe étant moins traitée que les autres en terme de temps, de rang, et avec des plus petites amplitudes.

On ne distingue pas comme dans le cas de la paire 1 des zones sémantiques fortement attractive. On observe que la figure choisie semble être explorée avec des durées de fixations plus courtes que la figure rejetée, mais pas ni plus ni plus tôt que la figure rejetée. La consigne ne semble donc pas diriger spécifiquement l'exploration, où bien l'effet de la position des figures est si fort qu'il bruite complètement cet effet. Nous allons observer si la consigne niveau de gamme entraîne le même type de regroupements.

✓ Cas 207-paire1, consigne niveau de gamme

▪ Classification

Il existe un traitement symétrique des zones Volant, Compteur, Levier de vitesse. Les classes sont plutôt mixtes en termes d'origine des zones.

▪ Discrimination

L'axe F1 comporte 80% de l'inertie et l'axe F2 20%. Le groupe 1, (celui contenant notamment les volants), est un groupe de zones traitées plutôt tôt dans la séquence, et plutôt fréquemment. Ainsi, la zone pédalier de la figure rejetée, la console centrale de la figure choisie, et les volants, sont les zones fixées le plus tôt et le plus fréquemment.

Les classes 2 et 3 sont donc des zones traitées plus tardivement et moins fréquemment. Ce qui distingue la classe 3 de la classe 1 et 2, hormis les dimensions précitées, est la durée moyenne des fixations, plus longues pour la classe 3. On trouve notamment dans cette classe la console centrale, les leviers de vitesse, le pédalier de la figure choisie, la console centrale de la figure rejetée. Ce sont donc des zones pertinentes, dont l'intégration cognitive ne semble pas avoir nécessité de nombreuses re-fixations. Enfin, la classe 2, qui contient entre autre les compteurs, la planche de bord de la figure choisie, l'ordinateur de la figure rejetée, se discrimine de la classe 1 et 2 car étant composée de zones fixées par de courtes fixations, et au total ne recueillant que peu de temps d'observation.

L'analyse de ce cas permet de décomposer les zones d'intérêts en trois classes selon le traitement visuel dont elles bénéficient, mais on ne peut ici relier la pertinence de la zone avec la classe à laquelle elle appartient. On observe cependant ici aussi que le traitement des zones pertinentes est souvent effectué par paire.

✓ Cas 207-paire2, consigne niveau de gamme

▪ Classification

Les zones regroupées au sein d'une même classe sont les zones volants et pédalier, et sont toutes dans la classe 1. Les classes sont mixtes (en termes de zones provenant du produit choisi ou rejeté).

▪ Discrimination

La dimension F1 contient 80% de l'inertie et la dimension F2 en contient 20%. Ces deux dimensions permettent de discriminer les 3 classes définies.

La classe 1, contenant les volants et les pédaliers, contient des zones traitées plus souvent, cumulativement plus longuement, mais aussi avec des durées de fixations plus longues. Il s'agit donc de zones vers lesquelles l'attention s'est fortement orientée. La classe 3 contient des zones traitées tôt dans la séquence. On y trouve notamment le levier de vitesse de la figure rejetée, l'ordinateur et la planche de bord de la figure choisie. La classe 2 contient des zones traitées peu souvent, plutôt tardivement, et avec de faibles amplitudes de saccades. Hormis le compteur de la figure rejetée qui peut être une zone pertinente par comparaison, elle est composée de zones portant peu d'informations pour la consigne.

Dans ce cas, on observe donc 3 classes, l'une contenant des zones beaucoup traitées, une autre contenant des zones traitées tôt, et une autre contenant des zones vers lesquelles l'attention ne semble pas s'être orientée spécifiquement.

✓ Synthèse des résultats pour le produit 207

La tâche de comparaison par paire a vraisemblablement engendré une observation symétrique de la scène car de nombreuses zones « jumelles » sont classées dans le même groupe (donc bénéficiant du même traitement visuel). Or les zones qui sont regroupées sont souvent des zones que l'on peut

considérer comme pertinentes pour l'évaluation. Cette méthode peut donc permettre de mettre en évidence une partie des zones d'intérêts pertinentes en observant celles qui sont classées dans le même groupe comportemental. Notons cependant qu'il n'y a pas d'effets notables de la consigne ; la reconnaissance de zones 'pertinentes' pourrait donc être d'une autre nature (distinctivité, saillance, complexité visuelle, informations similaires pertinentes pour la consigne, etc.).

La composition des classes semble posséder une certaine structure selon les différentes conditions expérimentales :

- existence d'une classe contenant des zones traitées plus tôt, plus souvent, plus longuement (zones « principales »). On y trouve les volants, mais aussi d'autres zones pertinentes telles que compteur ou levier de vitesse. Ce groupe serait celui des zones attirant spécifiquement l'attention. L'attractivité peut être due à plusieurs raisons : saillance, prégnance, pertinence, centre de gravité. Toujours est-il que l'on identifie par ce moyen un groupe de zones traitées tôt dans la séquence temporelle mais aussi de manière récurrente.
- existence de deux classes 'secondaires' (traitées moins rapidement, moins longuement, moins fréquemment que les zones « principales ») composées de zones traitées tôt dans la séquence. Dans ces classes secondaires on trouve une classe de zones plutôt « sémantiquement discriminées au regard de la consigne » et une classe de zones moins « pertinentes ». La classe de zones « pertinentes » se distingue car associée à des fixations aux durées et amplitudes plus longues que celles de l'autre classe secondaire.

Pour le produit 207, cette méthode ne met pas en évidence de stratégies de prise d'information propres à une consigne ou à une modalité visuelle, car on retrouve peu de points communs dans la composition des classes par les zones d'intérêt ni dans les caractéristiques temporelles du traitement visuel des zones. L'effet de la consigne et de l'interversion de la position des deux modalités visuelles a pu brouter les effets spécifiques de la consigne si tant est qu'ils existaient. Cependant, même si le principal facteur d'orientation de l'attention est de nature exogène (saillance physique) la consigne peut avoir un effet sur la dimension temporelle du traitement des zones, et donc permettre de dégager, sur la base des variables temporelles, des zones communément traitées visuellement dans les différents essais.

L'étude du produit C6 va permettre de confirmer si la tendance de classification des zones en trois groupes se confirme.

✓ *Cas C6-paire 1, consigne raffinement*

▪ *Classification*

Seules les zones 'console centrale' sont appareillées et qui, avec la zone 'volant' de la figure rejetée, constituent la classe 2. La classe 3 inclue des éléments centraux de la figure choisie, et de deux éléments centraux de la figure rejetée. La classe 1 est composée à part égale de zones de chaque figure.

▪ *Discrimination*

La dimension 1 (75% de l'inertie) est expliquée principalement par la durée des fixations. La dimension 2 (25%) est plutôt corrélée à l'amplitude moyenne de la saccade précédente.

La classe 2 se distingue des deux autres car ses zones sont traitées plus tôt, plus souvent, plus longtemps en cumul, que celles des classes 1 et 2. Les deux consoles centrales de la figure choisie constituent cette classe ; de plus le volant de la figure rejetée a été reclassé par l'A.F.D. dans cette classe. On y trouve donc en plus des consoles centrales, les deux volants. Les deux autres classes correspondraient à des zones traitées de manière secondaire, c'est-à-dire moins souvent, moins longuement en cumul, et plus tardivement. La classe 1 se distingue de la classe 2 car contient des zones traitées avec des amplitudes de saccades plus faibles que celles de la classe

2, ainsi qu'avec des fixations de durée plus courte. Ainsi, on constate que les éléments centraux qui constituent la classe 3 sont comparés par de grandes amplitudes de saccades et de longues fixations, tandis que les éléments de la classe 1 sont des zones peu traitées.

Il apparaît donc ici de nouveau une catégorisation en 3 types de zones : les zones principales, les zones secondaires attractives, les zones secondaires peu traitées. La figure choisie (classe 3) est donc traitée avec des fixations plus longues et ces zones attirent le regard « de loin », d'où de grandes amplitudes. Cela pourrait-il signifier que le regard se dirige dans cette situation vers des zones précises tandis que l'exploration des zones de la classe 1 se fait plutôt par proches voisins sans but précis ?

Nous allons confirmer ces observations avec l'étude des autres cas du produit C6.

✓ Cas C6-paire 2, consigne raffinement

▪ Classification

4 classes sont définies. La classe 1 est principalement composée de zones de la figure rejetée. La classe 2 est composée du volant de la figure rejetée et de la planche de bord de la figure choisie. La classe 3 est composée du volant et de la console centrale de la figure choisie. La classe 4 est composée d'éléments de la figure choisie et de la figure rejetée. Notons enfin que les zones compteurs, rétroviseurs, aérateurs centraux, et commandes du volant sont traités symétriquement (paires dans la même classe).

▪ Discrimination

3 dimensions sont nécessaires pour discriminer les 4 classes. La dimension 1 (53% de l'inertie) est décrite par le temps cumulé par zone et par la fréquence de traitement. La dimension 2 (36%) est associée au rang de Bradley-Terry, c'est-à-dire au traitement temporel des zones. La dimension 3 (10%) est définie par l'amplitude moyenne des fixations et la durée des fixations.

Les 4 classes se discriminent de la manière suivante. La classe 1 contient des zones traitées plutôt peu (ni très souvent ni très longuement au cumul), avec des amplitudes plutôt fortes et des fixations plutôt courtes. La classe 2 contient des zones traitées très tôt et plutôt régulièrement. La classe 3 contient des zones traitées avec de longues durées de fixations et plutôt souvent. La classe 4 contient des zones traitées tardivement, pas très souvent, avec des durées de fixation très variables mais des amplitudes de saccade plutôt faibles.

On observe donc que les zones de la figure rejetée ne sont pas beaucoup traitées, mais le sont avec des saccades de grande taille, sans que le regard ne s'y attarde. Le volant de la figure rejetée et la planche de bord de la figure choisie sont traités en premier et régulièrement ; ce sont des zones centrales spatialement, étant donc des centres de gravité, ou de stabilisation du regard. Le volant et la console centrale de la figure choisie sont traitées avec de longues durées de fixation et régulièrement ; ces zones peuvent aussi être des centres de stabilisation du regard. Enfin, il existe un groupe de zones secondaires, traitées peu souvent et plutôt par proche voisins.

On distingue donc :

- les zones fixées tôt dans la séquence et les zones fixées avec de longues fixations, (classe 2 et 3), qui sont des zones de stabilisation du regard
- les zones fixées peu souvent mais avec des grandes amplitudes et faibles durées de fixations, qui peuvent être des zones sur lesquelles s'effectuent une vérification comparative
- les zones secondaires qui ne se démarquent pas, plutôt traitées par proche voisins.

On constate comme pour le produit 207, que la consigne n'a pas structuré spécifiquement la sélection visuelle de zones d'intérêts. On retrouve comme pour la paire 1 l'idée qu'il existe un groupe de zones principales, dont l'attractivité peut être de nature spatiale, sémantique, ou encore de saillance ; il existe ensuite un groupe de zones secondaires traitées de manière plutôt volontaire et comparative, et enfin un groupe de zones secondaires traitées peu et plutôt par proches voisins. Cependant, ce ne sont pas les mêmes variables qui discriminent les groupes secondaires par rapport à la paire 1 malgré le fait que les zones de stabilité du regard sont les mêmes, à savoir volant et console centrale.

✓ Cas C6-paire 1, consigne niveau de gamme

▪ Classification

4 classes ont été définies. La classe 1 contient notamment les paires de zones suivantes : ordinateur, aérateur central, et levier de vitesse. Elle est équilibrée en termes d'origine de zones des figures 1 et 2. La classe 2 contient entre autre les paires 'logo, 'boite à gants'. La classe 3 contient la paire de zones 'console centrale' ainsi que le volant de la figure choisie. La classe 4 contient les commandes volant de la figure choisie et la zone centrale de la scène. On constate donc qu'il y a plusieurs zones regroupées par paire dans une même classe.

- *Discrimination*

La dimension 1 (64% de l'inertie) est définie par les variables : rang de Bradley-Terry, fréquence, temps cumulé. La dimension 2 (30%) est associée aux variables durée moyennes des fixations et amplitudes de saccades. La dimension 3 (7%) est associée aux durées moyennes des fixations.

La classe 3 (zones console centrale ainsi que volant de la figure choisie) se discrimine très fortement selon la dimension 1, donc contenant des zones traitées, à la fois très tôt et très souvent. La classe 4 (commande volant et centre scène) contient des zones très peu traitées et très tardivement, par des saccades de petite amplitude. Les classes 1 et 2 se discriminent car la classe 1 serait traitée avec des fixations plus courtes et des amplitudes plus grandes que celles de la classe 2.

On retrouve donc ici aussi un groupe principal composé des zones volant et console centrale, puis des groupes secondaires. Un groupe secondaire est négligeable : la classe 4, car les zones sont quasiment ignorées. Un groupe secondaire est traitée plutôt par proche voisins : la classe 2, et un groupe traité plutôt par comparaison, la classe 1, contenant notamment les paires d'ordinateur, aérateur central, et levier de vitesse. Il semble difficile de définir si l'attention s'est plus portée sur l'une de ces deux classes.

Cette consigne d'évaluation du caractère haut de gamme ne semble pas engendrer la sélection de zones spécifiques hormis les zones centrales, mais ne semble pas non plus délaisser certaines zones (hormis la classe 4). Y aurait-il pour cette consigne une exploration globale, rayonnant autour de centres de gravité, tandis que la consigne d'évaluation du caractère raffiné éliminerait un certain nombre de zones secondaires ? Cette hypothèse va être vérifiée sur la paire 2.

✓ *Cas C6-paire 2, consigne niveau de gamme*

- *Classification*

3 classes sont définies. La classe 3 contient notamment les paires de 'volant', et 'console centrale', ainsi que la planche de bord de la figure rejetée. La classe 2 contient entre autres les paires 'aérateur de droite', 'levier de vitesse', 'ordinateur'. La classe 1 contient des zones de la figure choisie ainsi que la zone 'commandes au volant et 'cave à pieds droite' de la figure rejetée.

- *Discrimination*

La dimension 1 contient 85% de l'inertie et la dimension 2 15%. La première dimension est associée au temps cumulé par zone et fréquence de traitement, tandis que la deuxième dimension est associée à la durée moyenne des fixations et l'amplitude de saccade.

La classe 3, qui contient les paires de volant et de console centrale, est traitée à la fois plus tôt, plus souvent, et plus longuement que les deux autres classes.

Les classes 1 et 2 se discriminent sur la dimension 2, car la classe 1 est traitée avec des amplitudes plus grandes et des durées plus longues que la classe 2. La classe 1 est aussi traitée un peu plus tardivement que la classe 2. La classe 1 contient plutôt des éléments locaux de la figure choisie, probablement traités de manière comparative, pour une vérification visuelle tardive par exemple. La classe 2 contient des éléments explorés par proches voisins, et plutôt longuement, donc correspondant plutôt à une investigation détaillée de chaque produit.

Il y aurait donc 3 phases correspondant à 3 groupes : fixation des zones centrales, qui seront ensuite des zones de stabilisation du regard ; puis, observation en détail et rapide de chaque figure (levier de vitesse, ordinateur, ...), et enfin confirmation de certains éléments sur la figure choisie (compteurs, cicle, rétroviseur,...).

✓ *Synthèse pour le produit C6*

Les zones volant et console centrales sont pour toutes les conditions expérimentales des zones qui bénéficient d'un traitement visuel appuyé, et vers lesquelles le regard se dirige rapidement. Les raisons pour lesquelles ces zones se distinguent peuvent être multiples : centres de gravités vers et depuis lesquelles le regard rayonne, saillance physique, zones portant l'identité de la scène (un

habitable est nécessairement composé d'un volant). Deux types de stratégies de prise d'information semblent se dégager : exploration par proches voisins et exploration par comparaison, sans qu'il ne se dégage de stratégies de sélection d'informations visuelles par l'orientation du regard spécifiques à une consigne.

✓ *Synthèse de la méthode*

Pour les quatre conditions expérimentales sur le produit 207, la variable durée des fixations est orthogonale à la variable « score de Bradley-Terry » des rangs de fixations. Cela signifie que la distribution des durées des fixations ne suit pas de tendance temporelle. Il n'y a que très peu de points communs dans la composition des groupes de zones entre les différents cas étudiés, hormis le fait que le volant de la modalité visuelle gauche et la console centrale de la modalité visuelle choisie font partie des zones observées le plus tôt et le plus souvent.

Pour C6, on observe de manière commune aux quatre cas une prise d'information préférentielle depuis des centres de gravité visuels. Les zones 'volant' et 'console centrale' sont ainsi les zones les plus traitées en durée mais aussi le plus tôt dans la séquence. Viennent ensuite des zones « satellites », dont la valeur sémantique n'est pas nulle mais secondaire, et enfin des zones qui semblent être explorées sans motivation particulière.

Il ne ressort pas de cette étude qu'un facteur particulier (consigne, modalité visuelle, etc.) oriente le comportement du regard de manière homogène pour les différents essais. L'inférence des informations utilisées pour le jugement à partir des positions du regard est ainsi limitée, notamment car :

- le participant peut ne construire son jugement qu'en fin de séquence sur la base d'informations encodées en mémoire et donc ne pas rechercher activement d'informations visuelles spécifiques lors de la prise d'information,
- la densité des informations dans l'image ne nécessite pas nécessairement une prise d'information locale, c'est-à-dire que le coût de déplacement oculomoteur du regard est jugé trop important par rapport au gain d'information pour modifier la position de prise d'information.

De plus on peut dégager au moins 3 raisons pour lesquelles une zone supporte des fixations oculaires hormis le processus de jugement :

- attractivité intrinsèque de la zone – distinctivité
- attractivité du fait d'un processus comparatif – effet mémoire dû au prototype cognitif
- attractivité spatiale du fait du traitement d'un proche voisin

Cette méthode met en évidence deux types de stratégies d'exploration de la scène (proche voisin, comparative), ainsi que différentes manifestations de l'attractivité des zones visuelles (traitées tôt/tardivement, longuement/rapidement, depuis un point éloigné, depuis un point voisin).

De ces résultats découle une meilleure compréhension des différentes stratégies de prise d'information visuelle dans une tâche de comparaison par paire, sur la base de classifications de zones visuelles, sans que l'on puisse relier ces analyses au traitement des informations visuelles. Enfin, le cas échéant, cette méthode permet la confirmation d'hypothèses spécifiques sur la pertinence de zones pour la consigne.

7.2.6 Existence de stratégies globales interindividuelles

Afin d'observer s'il existe des groupes de stratégies de prise d'information visuelle sur la base du comportement du regard, on relève pour chaque essai un ensemble de variables permettant de

caractériser temporellement et spatialement la prise d'information visuelle de chaque participant. Une Analyse en Composantes Principales de ces variables sur tous les essais est appliquée pour dégager des groupes d'essais dont le comportement oculomoteur est similaire. Les variables choisies sont les suivantes :

- moyenne des durées de fixation et nombre de fixations dans la séquence : exploration visuelle rapide ou lente
- cumul et moyenne des amplitudes de saccades : exploration visuelle étendue ou localisée
- moyenne des densités sur les zones fixées : exploration visuelle hétérogène ou homogène par rapport aux autres essais
- nombre de zones d'intérêts traitées : attention visuelle plutôt locale ou globale

La figure 7-11 représente les différentes dimensions de l'A.C.P. avec le positionnement de chaque individu.

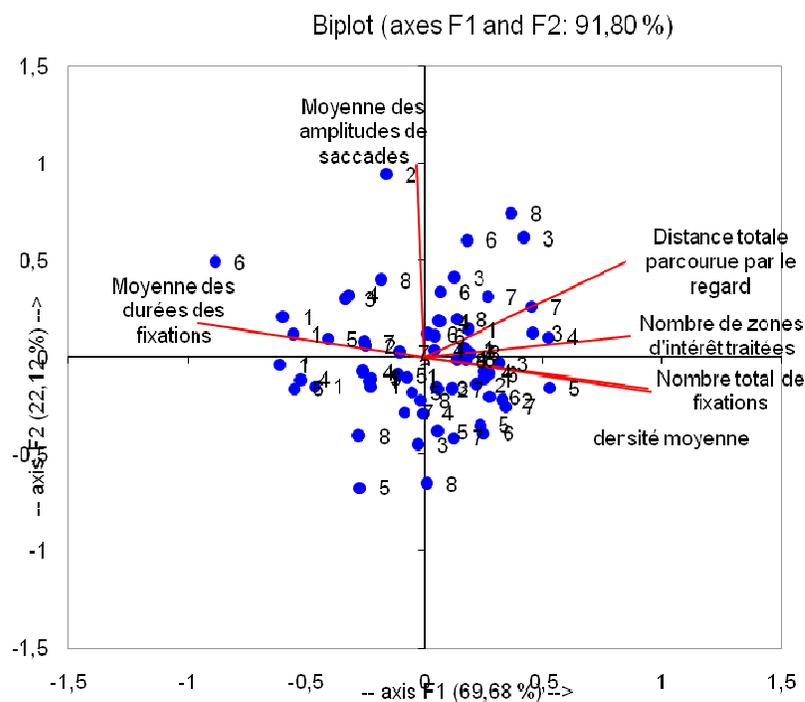


Figure 50 Analyse en Composantes Principales des essais en fonction des caractéristiques oculomotrices.

84% de l'inertie est représentée par deux dimensions (Figure 50). On constate que le positionnement des essais ne permet pas d'établir des groupes distincts d'essais selon leur caractérisation oculomotrice. La prise d'information sur la scène sur ce type de paradigme ne semble pas refléter des types de stratégies bien distincts.

7.2.7 Confrontation de la verbalisation à la prise d'information

Un tableau synthétique des observations verbales est représenté en Annexe C. A l'issue de chaque essai, le participant spécifie les zones visuelles qu'il considère décisives pour le choix. On étudie ainsi

s'il y a un lien entre les critères décisionnels exprimés post-exposition par les participants, les hypothèses des concepteurs et la durée d'observation de chaque zone.

Pour le produit 207, quelle que soit la consigne, les zones citées sont similaires. Il s'agit des zones qui diffèrent fortement entre les deux modalités, et qui correspondent à celles définies par les concepteurs comme portant sémantiquement le « typage ». Cependant l'interprétation est limitée par le fait qu'il peut s'agir de zones de décision, mais aussi des zones qui diffèrent visuellement entre les deux modalités, ou encore des zones qui diffèrent avec le prototype mnésique de l'objet.

Pour C6, la différence entre les deux produits est la couleur dominante blanche ou noire, et il n'y a pas de zones particulièrement évoquées par les sujets hormis la couleur. La couleur est ainsi l'élément le plus cité par les sujets. Les zones définies par les concepteurs sont peu citées car liées aux textures, attributs peu évoqués par les participants. L'évaluation du niveau de raffinement et du niveau de gamme entraîne les mêmes citations.

Au niveau de la relation entre le temps passé par zone et sa fréquence de citation, il apparaît que le temps passé par zone n'est pas directement corrélé avec la fréquence de citation. Ainsi, pour 207, des zones considérées comme importantes (haute fréquence de citation), tel que pédalier, levier de vitesse, sont observées avec des durées cumulées peu élevées, tandis que des zones fixées longuement (console centrale, ordinateur) sont citées par moins de 50% des participants.

Pour C6, la principale différence entre les deux modalités visuelles étant la couleur, c'est cette propriété qui est principalement citée. Les autres zones citées sont situées au centre des modalités, mais moins de 50% des participants les citent.

Au niveau de la relation entre le temps passé visuellement sur une zone et l'importance informationnelle qu'on lui accorde *a posteriori*, il n'apparaît pas de corrélation : des zones peuvent être observées très brièvement tout en étant très souvent citées, et des zones longuement observées ne sont pas nécessairement considérées comme importantes pour l'évaluation. L'étude essai par essai de la relation entre les zones citées et le temps cumulé de la zone ne montre non plus aucun lien spécifique.

Ces observations seraient expliquées par les hypothèses suivantes :

- la relation entre le temps cumulé par zone et le poids de la zone dans la décision est une hypothèse qui ne se vérifie pas
- la densité visuelle des objets dans la scène est telle qu'il n'est pas nécessaire de placer précisément son regard sur un objet pour le traiter visuellement, dé-corrélant position du regard et attention visuelle
- il n'est pas établi que les critères de décision soient définis au début ou pendant l'exploration visuelle de la scène, le jugement étant alors dé-corrélé temporellement de la prise d'information
- du fait de la justification *a posteriori* des critères de décision, et du fait des nombreux biais des processus de décision, les critères de décisions évoqués ne sont pas ceux qui ont été établis pendant l'exploration visuelle

7.3 Conclusion

Cette étude par comparaison par paire est intéressante dans la mesure où les processus de prise d'information sont moins statiques que dans le cas de la présentation monadique. On constate que les évaluations suivent les mêmes tendances (les mêmes produits dans les deux situations étant considérés les plus sportif, haut de gamme et raffinés) que dans le cas monadique. Cependant, quel que soit le type de traitement, il ne semble pas évident de corrélérer la prise d'information locale et l'évaluation. On note malgré tout que le produit choisi semble être observé plus longuement et plus tôt que le produit rejeté. Il serait donc intéressant de vérifier l'hypothèse qu'un pré-choix soit établi depuis un point de vue central. Cette méthode permet de mettre en évidence des tendances de stratégies exploratoires selon la modalité visuelle ou la consigne, non pas par la mise en évidence de zones visuelles spécifiquement traitées dans chaque cas, mais plutôt par le nombre de zones traitées, par le traitement ou non de zones jumelles, par la succession de processus comparatifs et analytiques. Ce paradigme permet aussi d'observer les liens entre les différents descripteurs du comportement oculomoteur du fait de la dynamique de l'exploration visuelle. Cependant, on note un fort biais d'observation des deux modalités visuelle de la paire depuis un point de vue central à chacune, et il n'apparaît pas de lien direct entre les zones verbalisées comme décisives et le comportement oculomoteur. A la suite de ces deux expérimentations d'évaluation monadique et par paire, il semble donc que l'orientation locale du regard soit dissociée du processus de jugement, et que le nombre de facteurs qui guide le regard ne permette pas une forte homogénéité de traitement des informations visuelles propres à chaque condition expérimentale. S'il semble possible de mettre en évidence des phases cognitives temporelles qui se reflètent sur le comportement du regard, il semble difficile d'associer les informations visuelles fixées à un facteur attentionnel spécifique. Si le système visuel est effectivement soumis aux demandes attentionnelles pour fournir des informations visuelles précises via le repositionnement du regard, la complexité des interactions entre les processus activés ainsi que la densité des informations visuelles dans la scène expliquent la difficulté de faire émerger les informations visuelles traitées spécifiquement aux tâches évaluatives.

A la suite de ces observations, plusieurs processus complémentaires peuvent être étudiés. Lors de l'expérimentation monadique, l'hypothèse a été soulevée que le regard puisse être soumis à un biais d'observation centrale, inhibant les mouvements du regard vers des zones locales spécifiques. Dans le paradigme d'exposition par paire, l'existence d'un processus d'orientation rapide du regard par un pré-choix a été évoquée, où le regard se positionne dès la première seconde sur l'élément qui sera choisi. Il semble ainsi pertinent de tester la qualité des informations visuelles qui peuvent être extraites depuis un point de vue central fixe de la scène. Si l'ensemble des informations visuelles nécessaires à la consigne peuvent être extraites depuis un seul point de vue, il est alors possible de confirmer l'hypothèse d'une pré-décision dans la première seconde d'exposition, mais il devient très difficile de corrélérer les lieux de prise d'information visuelle avec les informations effectivement traitées par le système cognitif. Cette étude est mise en place sous la forme d'un paradigme de vision contrainte, dont les résultats sont présentés dans le chapitre suivant.

La deuxième étude que l'on effectue porte sur l'interaction entre la position du regard, l'attitude (au sens psychologique du terme, explicité en partie théorique) et la décision. Il a été observé dans la littérature que le regard obéissait non seulement à une fonction de capture d'informations visuelles, mais aussi à une fonction de manifestation de l'attitude et d'aide à la prise de décision. Plutôt que d'observer les zones locales sur lesquelles le regard s'est posé, on va chercher à identifier sur la base du comportement oculaire un processus d'orientation globale et non locale du regard. Si la prise d'information visuelle est suffisamment efficace depuis un point de vue central, la position du regard

n'est pas informative de la sélection d'informations visuelles, mais peut l'être en termes d'attractivité globale d'une des modalités visuelles. Dans les paradigmes d'évaluation monadique et par paire, la durée d'exposition était fixe, contraignant le participant à observer l'image même si son choix était effectué. Dans cette présente étude, chaque juge mettra fin à l'essai lorsqu'il aura terminé la tâche, afin de synchroniser les processus de prise de décision entre participants.

La troisième étude complémentaire consiste à établir un modèle d'évaluation de la contribution relative de différents facteurs de guidage du regard évoqués dans notre travail. Pour ce faire, on effectuera tout d'abord une analyse des zones locales qui sont des critères de décision pour le jugement, puis une analyse de la qualité de prédiction du comportement visuel par les algorithmes de saillance. Enfin, nous présenterons une méthode basée sur les mélanges de gaussiennes pour évaluer quels facteurs attentionnels expliquent la distribution des points de fixation, et donc dans le cadre de nos paradigmes s'il peut y avoir des liens entre l'attention visuelle et le jugement.

Chapitre 8

Paradigme de vision contrainte

8.1 Matériel et méthodes

8.1.1 Présentation

L'étude du comportement visuel en situation d'évaluation de produits manufacturés a mis en évidence l'existence de zones visuelles vers lesquelles l'attention se porte ainsi que différentes stratégies d'exploration de la scène. Or il n'a pas été observé de lien direct entre le processus de jugement et la prise d'information locale. Du fait de la densité visuelle des objets composant la scène, de l'existence de point de vue optimisant la prise d'information, et de la pré-connaissance par les participants de ce type d'objets, de nombreuses informations ont pu être traitées sans avoir été fixées. Dans une situation d'évaluation monadique, il existe un fort biais de centralité (c'est-à-dire l'observation de la scène depuis une zone centrale). Dans la situation d'évaluation par paire du produit 207, la modalité visuelle qui est choisie à la fin de l'exposition semble être significativement fixée par les premières fixations, comme si un pré-choix avait pu être effectué très rapidement par le biais d'informations visuelles périphériques.

Cette étude a pour objectif de vérifier si depuis un point de vue central correspondant à la position initiale du regard, le participant est capable d'extraire suffisamment d'informations visuelles pour effectuer un pré-choix, expliquant les premiers mouvements du regard, et si les principaux objets de la scène peuvent être traités sans qu'il ne soit nécessaire de les fixer localement (expliquant le biais de centralité). Pour vérifier ces hypothèses, un paradigme de « vision contrainte » a été mis en place.

8.1.2 Description du paradigme

✓ Objectifs

Le paradigme de vision contrainte consiste à contraindre le participant à une observation de la scène visuelle depuis une zone définie *a priori* et uniquement dans cette zone. Ici nous avons choisi la zone centrale. Ainsi, dès que le regard sort de cette zone centrale, la scène visuelle disparaît. Seules les informations visuelles du centre de la scène sont donc fixées, les autres étant donc acquises par la vision périphérique.

Dans le cas de l'évaluation monadique, si l'habitable peut être précisément décrit alors que les informations excentrées ne sont pas fixées, cela confirme l'hypothèse que la position du regard n'est pas directement corrélée aux processus de sélection des informations visuelles, et que l'on peut envisager l'existence de centres de gravité depuis lesquels la prise d'information est optimisée. Les déplacements du regard reflètent alors des processus attentionnels et non pas la sélection des informations visuelles nécessaires à l'évaluation. Dans le cas de la comparaison par paire, si l'évaluation est possible depuis la zone centrale de la scène, et si les participants décrivent correctement les deux modalités visuelles, alors l'hypothèse qu'il n'est pas nécessaire d'explorer

localement les deux images pour faire un choix est confirmée, et donc qu'un pré-choix rapide et par la vision périphérique est envisageable.

✓ *Dispositif*

Un logiciel gérant l'affichage des images couplé à un eye-tracker « eyelink 2 » (mesure à 250hz : latence de 4ms) affiche la scène visuelle à l'écran si le regard est dans la zone centrale. Dans le cas contraire, l'image disparaît. L'image est affichée sur un écran cathodique de résolution 1024*768, d'une largeur de 40cm ; le participant est placé à 55cm de l'écran. Le visage est positionné sur une mentonnière afin de minimiser les mouvements de tête. Chaque essai débute par la calibration du système. Le participant est ensuite convié à positionner son regard sur un disque de couleur au centre de l'image, défini par une zone de 100*100px (diamètre de 4° d'angle visuel – environ 4 cm sur une largeur d'écran de 40 cm, soit 2*22° du champ visuel). On note que le ratio distance/taille d'image est identique à celui des expérimentations 1 et 2, c'est-à-dire que la taille angulaire de chaque objet sera similaire. On considèrera donc que les dimensions du champ visuel couvert par l'écran sont identiques. La stabilisation du regard dans la zone centrale (120*140px, environ 5°) pendant plus de 60ms déclenche l'exposition de l'image, et ce tant que le regard reste dans cette zone. Un entraînement apparaît nécessaire pour réussir à inhiber les mouvements involontaires du regard hors de la zone centrale ainsi que pour diminuer le coût attentionnel de la tâche dû à la contrainte imposée. Chaque participant effectue 10 essais (6 images de paires, 4 images monadiques) dont l'ordre est défini aléatoirement.

✓ *Produits*

Puisque l'on souhaite vérifier des hypothèses portant sur l'évaluation des produits de l'expérimentation 1 et 2, on réemploie ici les mêmes images de C6 et 207, tant en présentation par paire que monadique. Afin d'enrichir les observations, on va employer des images complémentaires, représentant les habitacles des modèles Lancia Upsilon et Mini Cooper. Ces modèles sont choisis car étant de la même gamme que la 207. L'Upsilon est un produit plutôt classique, avec cependant un positionnement des compteurs en zone centrale, donc une modification de la configuration classique de scène. La Mini-Cooper est choisie car étant très différente en terme d'évocation sémantique, de positionnement des objets, et aussi de taille des objets, que les autres habitacles étudiés ; deux versions sont exploitées (voir Annexe 3).

✓ *Consignes*

Les deux types d'évaluation précédemment employés (monadique et par paire) sont testés : évaluation de la sportivité, du niveau de gamme, et pour C6 du caractère raffiné. Dans le cas de l'évaluation monadique, il est demandé au participant de relever pour un test de mémoire le maximum d'informations permettant de décrire le produit, en le verbalisant en temps réel. Dans le cas de l'évaluation par paire, les participants doivent évaluer lequel des deux habitacles est le plus sportif, tout en décrivant à voix haute la scène perçue. Il est spécifié aux participants de maintenir leur regard dans la zone centrale, sans quoi l'image disparaîtra.

✓ *Recueil*

L'approche qualitative qui est adoptée ici se base sur les relevés des verbalisations et le questionnement de douze participants âgés de 24 à 29 ans (m=27,2), tous doctorants au laboratoire LIS/INPG et ayant une vision normale. Le type de population est certes différent de celui des expérimentations 1 et 2 mais puisqu'il s'agit d'un test de nature plutôt psychophysique qu'évaluatif, on considèrera cette différence de population sans conséquences sur les résultats. Les relevés sont les descriptions verbales de la scène pendant la visualisation. Afin de vérifier que les objets excentrés sont bien traités, si ceux-ci n'ont pas été librement évoqués, l'expérimentateur en demande la description, tels que la forme du rétroviseur, les matériaux des aérateurs latéraux et du pédalier. Le

volume et la gamme du véhicule est demandé, tant pour la modalité d'exposition par paire que monadique. Enfin, les jugements et les critères de décisions sont relevés.

8.2 Observations

8.2.1 Effets de la contrainte de la position du regard

Chaque participant effectue un entraînement sur des photographies de scènes réelles jusqu'à ce qu'il soit capable de maintenir volontairement son regard dans la zone centrale : le premier entraînement est difficile pour tous les participants, le regard quittant la zone centrale en moins d'une demi seconde. Le deuxième entraînement est de meilleure qualité, le regard étant bien stabilisé sur la zone centrale, et l'image restant ainsi à l'écran plus d'une seconde. Le troisième entraînement montre que le système est plutôt bien manipulé, 9 participants contrôlant correctement la durée d'exposition de l'image.

Tous les participants considèrent que le maintien du regard sur la zone centrale est un effort attentionnel coûteux, c'est-à-dire que cela nécessite une attention volontaire. Un test complémentaire sur 3 participants montre qu'à l'issue de douze essais effectués en deux séquences espacés de 10 minutes, il est possible de contrôler la position du regard tout en effectuant des opérations de calcul mental, donc en réallouant l'attention hors de l'activité visuelle. La charge attentionnelle du maintien du regard est donc dépendante de l'entraînement.

Puisque l'on souhaite rester dans un contexte le plus proche possible des expérimentations précédentes malgré la contrainte, l'entraînement approfondi n'a pas été retenu. S'il permet un contrôle volontaire du regard, il produit aussi une optimisation des processus de discrimination visuelle et donc permet au participant entraîné d'extraire plus d'informations visuelles depuis les zones périphériques de l'image que le participant non entraîné.

8.2.2 Evaluation monadique

Tous les participants sont capables (soit d'eux même soit sur sollicitation) de décrire l'ensemble les objets en présence dans l'image alors même que le regard est orienté vers la console centrale lorsqu'il est posé au centre de l'image. La forme et couleur des compteurs sont évoqués spontanément dans 70% des essais, l'aspect métallique et les perforations des pédales dans 60%. La forme de la planche de bord (prototypique *versus* non prototypique) est décrite dans 50% des cas, les matériaux revêtant les sièges dans 70%. Si la composition ou la forme de la console centrale est non prototypique, elle est alors mentionnée et décrite comme « originale » (ex. avec la photographie de Mini Cooper)²². Tous les participants répondent correctement à la description des éléments tests (forme du rétroviseur, matériaux des aérateurs latéraux et pédalier). Quant aux informations de volume de l'habitacle (associée à la taille et la gamme dans les verbalisations), elles sont mal évaluées. Dans 50% des cas, il n'y a pas de réponse, et dans 25% l'estimation du volume est fautive. Enfin, l'évaluation post exposition est possible dans 80% des essais, et les réponses se classent en trois catégories : moderne ou classique (« trop ou pas assez »), toc ou sympa, jeune ou vieux.

²² Voir photographies en annexe

L'hypothèse que le positionnement central du regard permette d'extraire des informations sémantiques suffisamment détaillé des zones périphériques est donc confirmée. Il est ainsi probable qu'il existe des centres de gravité depuis lesquelles la prise d'information est optimale, et donc qu'il n'est pas nécessaire de positionner le regard localement sur un objet visuel pour le traiter sémantiquement.

8.2.3 Présentation par paire : évaluation de la sportivité

✓ *Une configuration de scène dissymétrique*

Les images présentées par paires sont plus complexes à traiter que les images en présentation monadique. En effet, la largeur du produit est diminuée par 2 par rapport à la taille d'exposition en monadique, le nombre d'information dans la scène est multiplié par deux, puisqu'il y a deux produits, et les informations sont spatialement plus étendues. Il existe de plus un déséquilibre de symétrie : s'agissant d'habitacles « français », le poste de conduite est sur la moitié gauche de chaque produit. Le regard étant contraint au centre de la scène, donc entre les deux images, il ne se trouve pas à équidistance des deux postes de conduites mais beaucoup plus proche de celui de droite. Afin d'y remédier, il eut été possible de présenter des habitacles antisymétriques (poste de conduite français vs. anglais), ou encore de les présenter verticalement. Mais cette expérimentation doit permettre de contrôler des hypothèses issues de l'expérimentation 2, il n'est donc pas pertinent de modifier ici la position des images dans la scène.

Les résultats montrent que les participants décrivent beaucoup plus aisément le poste de conduite du véhicule de droite que celui du véhicule de gauche. Ainsi, dans 70% des essais, les participants expriment le fait qu'ils sont gênés par le déséquilibre de configuration spatiale de la scène ou que la reconnaissance de l'image de gauche demande un effort attentionnel conséquent. Ainsi, la description de la scène passe d'abord par la description du produit de droite, beaucoup plus détaillée que celle du produit de gauche.

✓ *Choix de l'habitacle le plus sportif*

Alors même que les participants sont capable de maintenir leur regard volontairement, les essais sont classés en deux types : les essais brefs (inf. à 1s) et les essais longs (sup. à 2s). Les essais de durée intermédiaire étant quasi absents.

▪ *Essais courts*

Dans le cas des essais brefs (environ 30%, et principalement sur des paires composées de modèles différents), les participants évoquent *a posteriori* qu'un processus attentionnel involontaire et immédiat a attiré leur regard vers une zone non centrale (provoquant la disparition de l'image). Les participants évoquent principalement que le mouvement a eu lieu car l'une des deux images serait une représentation non classique de l'habitacle (ex : 207/mini). Les évocations sont de l'ordre de « l'anomalie », « la bizarrerie », « l'originalité ». On note donc une manifestation du schéma de scène (information mnésique) sur le guidage attentionnel. Aucun participant ne cite spontanément le modèle d'un produit, ils ne sont donc probablement pas reconnus. Pour plus de 90% des essais, les caractéristiques de luminance et de chrominance de chacun des deux produits sont correctement décrites. Le cas échéant, 80% des verbalisations citent la présence de formes « originales » *versus* « classiques ». Le choix de l'habitacle le plus sportif semble se baser sur des informations de luminance : plus il existe de zones brillantes sur un produit, plus il sera considéré sportif (70% des choix de sportivité des essais courts se basent sur ce critère). Ainsi, la décision, à défaut de se baser sur une réelle identification des images et des objets qui constituent l'habitacle, semble se reposer sur un critère de luminance et de contrastes ; il y aurait donc une orientation de l'attention visuelle par des facteurs physiques de nature « signal » convergeant avec des propriétés sémantiques.

- *Essais longs*

Dans le cas des essais longs (70%), les participants stabilisent leur regard sur la zone centrale en dissociant attention visuelle et locus perceptif. Ils accumulent de l'information jusqu'à ne plus détecter d'éléments nouveaux, effectuent alors un mouvement volontaire du regard pour mettre fin à l'essai. Les informations qu'ils sont capables de retranscrire sont assez homogènes.

- Retranscription des informations visuelles traitées :

Plus de 90% des informations retranscrites « en temps réel » dans la description de la scène font partie d'un de ces types: chrominance, contraste, originalité des formes, originalité de la configuration. 60% de ces informations sont issues de comparaison entre les deux modalités visuelles (descriptions comparatives) et 40% par référence à la mémoire (descriptions non comparative). Ainsi, les différences de couleur et de luminance sont détectées et retranscrites. Cependant, malgré la détection d'une différence de forme ou de configuration entre les deux produits, moins de 30% des détections sont accompagnées d'une identification de la nature de la différence de forme entre les deux habitacles ou la nature de « l'anomalie configurale » de l'habitacle.

Pour moins de 10% des essais, une réponse de l'évaluation du volume et du niveau de gamme a été fournie, s'avérant de plus très approximative.

Il apparaît donc que les informations perceptives sont plutôt bien traitées, que les informations configurales sont acquises mais non identifiées, et que les informations de volume et de taille ne sont quasiment pas traitées.

Au niveau des informations sémantiques, les objets situés à proximité de la zone centrale de l'image sont identifiés, tels que le volant et le compteur de la figure de droite. Le pédalier de l'habitacle positionné du côté droit est correctement décrit par tous les participants si celui-ci possède des caractéristiques perceptives saillantes. Cependant, l'habitacle de gauche est très mal analysé en terme sémantique, décrit par les participants comme étant trop éloigné.

- Évaluation :

Les informations de contrastes sont rapidement acquises car sont décrites après un essai court. Or, près de 70% des évaluations de la sportivité prennent en compte les propriétés de luminance des habitacles. Ainsi, la figure « la plus lumineuse », ou comportant « le plus d'éléments brillants » est choisie dans plus de 80% des essais comme étant la plus sportive. Le produit considéré dans l'expérimentation 2 comme étant le plus sportif comporte effectivement plus de matériaux chromés, donc de zones brillantes. Les matériaux sont aisément et rapidement traités sans fixation et semblent inférer sémantiquement la sportivité. Ainsi, si la modalité visuelle de type « sport » est fixée en premier dans l'expérimentation 2, cela peut s'expliquer soit par une attractivité de nature saillante des zones de fortes luminances, soit par une association sémantique entre sportivité et brillance.

8.2.4 Synthèse

Cette étude dégage les informations suivantes. Dans le cas de la présentation monadique d'un habitacle, l'ensemble des informations présentes dans la scène peuvent être traitées depuis un point de vue central. La fixation d'une zone n'est donc pas nécessaire au traitement des informations visuelles locales qu'elle contient. Elle démontre plutôt la présence d'attracteurs attentionnels (de diverses natures) ou de l'existence d'un point de vue optimal. Le système visuel a ainsi, pour ce type d'images, à la distance employée, et pour une consigne de type évaluative, la capacité d'analyser l'ensemble des informations visuelles en présence dans l'image sans solliciter de mouvement du regard. Dans le cas par paire, sans même être capable de décrire complètement l'une ou l'autre des modalités visuelles, le participant détecte les dissimilarités ou les singularités. On appuie ainsi le

concept de « distinctivité » du stimulus, sur la base d'informations tant perceptives que cognitives. Cette propriété agirait en tant que facteur d'orientation de l'attention à un niveau global via un mécanisme d'attractivité et à un niveau local via un mécanisme de recueil d'informations. On note ensuite que les différences de luminance et de chrominance sont rapidement analysées et intégrées dans le processus de comparaison. L'organisation schématique de la scène et la forme des objets sont comparées, à la fois entre les deux modalités visuelles de la paire (référence), et à la fois entre une modalité visuelle et des informations mnésiques (prototype d'objet ou schéma de scène encodés en mémoire) afin de permettre la détection « d'anomalies » sans que cette dernière soit nécessairement identifiée.

Les informations traitées depuis le point de vue central de l'écran, à la distance utilisée dans les expérimentations monadique et par paire, et pour des tâches d'évaluation, sont ainsi premièrement perceptives, puis configurales²³, mais plus tardivement sémantiques. L'attention est donc principalement portée sur les caractéristiques physiques des modalités visuelles et sur les singularités configurales.

En reprenant les hypothèses, il est confirmé pour l'exposition monadique que le regard peut effectivement se positionner sur un point de vue central et traiter la globalité de la scène, mais pour l'exposition par paire, le point de vue central de l'image ne permet pas un traitement fin de chacune des deux modalités. Concernant l'évaluation des habitacles, il semble que l'exposition par paire ne permette pas de catégoriser réellement la sportivité au premier regard depuis un point de vue central, mais permet de détecter des matériaux qui peuvent activer par inférence la catégorie des attributs sportifs. Cela montre qu'il est nécessaire que le regard se repositionne sur chacune des deux images pour les analyser, ce qui se corrobore avec les observations de l'expérimentation monadique.

Ces éléments confirment de plus l'existence de multiples processus attentionnels :

- acquisition des informations globales perceptives selon leurs caractéristiques de saillance
- comparaison des informations perceptives avec les proto objets en mémoires, orientant le regard vers les éléments considérés comme « anormaux » (originalité de la structure ou de la configuration) ou « perceptivement perturbant » (forte luminance ou contraste élevé)
- attractivité sémantique lorsque la représentation d'un objet fait partie des attributs d'une catégorie

Cette expérimentation n'a cependant pas montré si en condition d'exploration non contrainte, le regard se pose naturellement sur les zones locales vers lesquelles s'oriente l'attention. Il n'y a sans doute pas de réponse unique, car selon le niveau de détail nécessaire à l'évaluation et l'évolution temporelle de la séquence, l'investigation sera plus ou moins locale et corrélée à la demande de tâche. Après avoir observé que la position du regard ne reflète pas directement la quantité d'informations encodées, et que la propriété de distinctivité de la modalité visuelle semble influencer l'orientation de l'attention, nous allons aborder la problématique de l'existence de liens entre l'attention, l'attractivité (l'attitude interne²⁴ produite par la première impression) et la décision, ainsi que l'existence d'une phase temporelle spécifique à la décision.

²³ Organisation de la scène en termes de position et de forme des objets

²⁴ Pour rappel, l'attitude interne est un état affectif, qui, porté sur un objet, est la conséquence d'une évaluation favorable ou défavorable.

Chapitre 9

Préférence et orientation du regard

9.1 Présentation

- *Contexte*

L'évaluation visuelle d'un produit active des mécanismes complexes d'acquisition et de traitement des informations visuelles afin de percevoir la scène, d'attribuer des significations, et enfin d'effectuer un jugement. Le regard agit comme un capteur permettant de recueillir des informations visuelles. Or, il s'avère que le regard peut aussi agir comme un pointeur comportemental de l'état cognitif et affectif. En effet, l'acquisition des informations visuelles peut être très rapide (moins d'une seconde pour l'identification par exemple), et le regard reflète ensuite des processus d'attractivité, de préférence et de décision sans fonction d'extraction de l'information. Différents travaux montrent que l'orientation du regard en direction d'un objet spécifique parmi d'autres démontre un intérêt, établissant un lien d'attractivité qui participe ensuite à la construction du jugement et à la prise de décision. Des études ont ainsi porté sur le lien entre l'orientation du regard et l'évaluation de visages mais aussi d'objets de la vie quotidienne (cravates, montres, boucles d'oreilles) [Simion05]. Cette interaction va se manifester dans deux sens. Tout d'abord, quel que soit le type d'évaluation comparative, le regard va manifester un processus d'attractivité puis de décision, en s'orientant de plus en plus préférentiellement vers le produit qui va être choisi, jusqu'à la décision. La deuxième manifestation de l'interaction regard-décision est dénommée « Gaze Cascad Effect » et consiste en un biais de décision lors de l'évaluation de la préférence. En effet, après émergence de l'attitude première envers chacun des deux produits, deux processus vont émerger. Le premier processus consiste à privilégier l'observation du produit envers lequel l'attitude est la plus positive. Le second est un processus produisant à juger plus favorablement le produit sur lequel on possède le plus d'informations, c'est-à-dire celui que l'on a le plus observé (aisance au traitement perceptif). Ainsi, plus je regarde un produit plus je l'apprécie, et plus j'apprécie un produit, plus je le regarde. Le temps passé à observer chaque produit va ainsi augmenter de plus en plus fortement en faveur du produit qui sera finalement choisi.

Le regard participe ainsi à la prise de décision, non pas uniquement en tant que capteur, mais aussi en tant que pointeur. Le fait de traiter facilement les informations visuelles d'une modalité visuelle, ou encore le fait d'avoir une première impression positive de cette modalité, va orienter le regard sans sa direction. Mais l'orientation du regard va marquer l'intérêt, va augmenter la connaissance de cet élément (d'où meilleure aisance de traitement), et va ainsi pointer la préférence. Les mouvements oculaires possèdent donc plusieurs fonctions, la première étant une fonction de réajustement du champ visuel pour l'extraction optimisée d'informations à hautes résolutions, la seconde étant une fonction de pointage comportemental d'états internes et ayant un rôle de manifestation de l'intérêt et d'orientation du jugement. L'étude du rôle de pointeur comportemental du regard complète notre travail car nos résultats précédents montrent que le regard est plutôt guidé par la nature perceptive de la scène et non par la demande cognitive. Il est ainsi intéressant d'observer si le processus cognitif de prise de décision va avoir ou non un effet sur le comportement oculomoteur. Il

est proposé dans la littérature que la consigne de préférence soit différente des autres évaluations en terme cognitif et de comportement oculomoteur. Dans le cadre de l'étude des objets manufacturés, il est pertinent de pouvoir confirmer si la consigne de préférence est bien une tâche cognitive spécifique (produisant un fort biais de type « cascaded effect »).

▪ *Observations*

Plusieurs hypothèses peuvent être formulées afin de répondre aux objectifs. La première hypothèse concerne l'effet de l'activité cognitive de décision sur l'orientation du regard et sera vérifiée par l'existence d'une phase de prise de décision distincte de la phase d'exploration. La deuxième hypothèse porte sur la spécificité de la consigne de préférence, et sera vérifiée par la comparaison du comportement oculomoteur selon les consignes sur la base de l'évolution temporelle du taux de fixation du produit choisi. De manière générale, nous comparons les taux de fixations sur la modalité visuelle choisie dans les secondes qui précèdent le choix (signalé par le participant), selon la difficulté de choix (modalités visuelles similaires ou non), et selon la consigne.

✓ *Dispositif expérimental*

▪ *Images*

6 paires de photos d'habitacles de véhicules sont construites. 3 paires représentent deux versions différentes d'un même habitacle (C6 et 207), 3 paires sont construites à partir d'habitacles de 3 véhicules (207-Sport/Mini ; Upsilon/207sport ; Mini/Upsilon ; cf. annexe 3). De par l'alternance de la position droite gauche, nous obtenons deux jeux de 6 paires. Pour la suite de cette étude, on notera « figure » la modalité visuelle représentant un habitacle.

▪ *Consignes*

4 consignes sont définies : évaluation de l'habitacle le plus sportif (« C1+ »), de l'habitacle le moins sportif (« C1- »), de l'habitacle préféré (« C2+ »), de l'habitacle le moins apprécié (« C2- »). Pour chaque évaluation, le participant est invité à répondre à la consigne en cliquant sur la flèche gauche ou droite du clavier dès que son choix est effectué. Il lui est précisé qu'il possède le temps qu'il veut pour répondre. Lorsque le participant a cliqué, la paire suivante apparaît, précédée par un écran noir et un pointeur central.

▪ *Séquence*

Une séquence consiste à évaluer 6 paires différentes dans un ordre aléatoire. Chaque participant effectue un entraînement sur deux images de peintures, la consigne étant de choisir l'image que l'on préfère. A la fin de la séquence, les participants sont de nouveau exposés aux paires afin de recueillir une note par produit, de 1 à 7 (de la moins xxx au plus xxx) ainsi que l'explicitation de leurs critères de décision.

▪ *Système*

Les paires sont présentées sur un écran de 167*124cm, à une résolution de 1400*1050, à une distance de 2,30m du participant, soit un champ visuel horizontal de 2*20° de part et d'autre de l'axe central de l'écran, dans des conditions identiques à celles des expérimentations précédentes. Les mouvements des yeux sont enregistrés avec le système Facelab® V4.3, à une fréquence de 60hz (soit un point tous les 17ms), en mode précision. La dérive moyenne constatée est d'environ 1°, soit environ 30px (2% de la largeur d'écran).

✓ *Méthodes d'analyse des données*

Le terme « figure » est employé pour définir chacune des deux modalités visuelles de la paire.

▪ *Synchronisation des données*

Tous les essais d'une même condition expérimentale sont combinés et alignés sur le moment de la décision. On observe ainsi les mesures les synchronisant sur le moment de la prise de décision et non sur le début de l'exposition. Ainsi, quelle que soit la durée de l'essai, l'observation des données porte sur les secondes qui précèdent le choix. Le taux d'impacts sur la figure choisie au cours du temps est donc établi jusqu'à la prise de décision. On définit comme « top d'horloge » chaque instant pour lequel ce taux est calculé, par regroupement de tous les points d'impacts mesurés entre ce top d'horloge et le top « n-1 ».

- *Comparaison au taux de hasard*

Pour chaque top d'horloge, un taux de hasard est calculé en établissant la valeur de la loi binomiale ($p=0,5$; confiance de 95%) au delà duquel le taux mesuré est au dessus du hasard, pour un nombre de points de mesures équivalent à celui utilisé pour calculer le taux instantané de points d'impacts sur la figure choisie. Si le taux d'impacts sur la figure choisi dépasse celui proposé par la loi binomiale, alors les regards sont significativement orientés vers l'une des deux figures. A l'inverse, si ce même taux est en deçà de la limite du taux de hasard, alors il n'existe pas d'orientation significative du regard.

- *Modéliser la distribution*

Nous reprenons la procédure qui est classiquement utilisée pour modéliser le taux d'impacts sur la figure choisie par une sigmoïde à 4 paramètres. Selon la littérature, les caractéristiques de cette courbe permet de distinguer les consignes par le comportement oculomoteur associé. Une forte courbure de cette fonction sigmoïde dans les instants précédant le choix est associé à la présence de l'effet « gaze cascade ». De ces résultats nous pouvons mieux comprendre les processus de décision pour chaque consigne, et les différences qui peuvent être imputées à la nature de la tâche.

- *Similarité des densités des fixations*

La distance de Kullback-Leibler fournit une distance entre deux distributions. Elle permet de classer les différentes fonctions de distribution des fixations sur la figure choisie et d'observer s'il existe un effet des consignes sur le processus d'orientation du regard.

- *Similarité non orientée entre distributions de fixations*

La distance de Mannan consiste à mesurer la similarité spatiale de deux distributions de fixations, afin de dégager si la prise d'information est explicable par des facteurs expérimentaux. Puisque cette distance n'est pas dépendante de l'orientation du nuage de points, il est possible de regrouper tous les essais effectués sur une paire quelle que soit la latéralisation relative des images de la paire. Une analyse de variance permet de comparer cet indice de proximité selon les facteurs expérimentaux qui auraient pu influencer la distribution des points de fixations.

- *Similarité intra condition expérimentale des essais*

La méthode de « Meanshift » (utilisé par Santella & Decarlo [Santella04]) consiste à regrouper les fixations en classes spatiales ('clusters'). L'idée principale est de pouvoir établir des zones d'intérêts *a posteriori* et non *a priori*, selon les zones définies par ces clusters. Chaque fixation est ainsi associée à un cluster, permettant une classification de ces zones spatiales en fonction du comportement oculomoteur. En appliquant un algorithme de type Levenstein (« string-edit ») ou de « data time warping », on dégage la similarité de fixation des zones entre séquences en prenant en compte l'ordre de fixation des zones et le nombre de zones fixées.

9.2 Résultats

9.2.1 Etude du temps nécessaire à la prise de décision

Pour chaque essai, le participant met fin lui-même à l'exposition lorsqu'il a terminé son jugement. Afin de vérifier si la nature de la consigne influe sur le délai de la prise de décision, une analyse de variance est effectuée sur le facteur consigne et la variable durée d'exposition : il n'y a pas d'effet significatif de la consigne ($p=0,074$) sur la durée des essais.

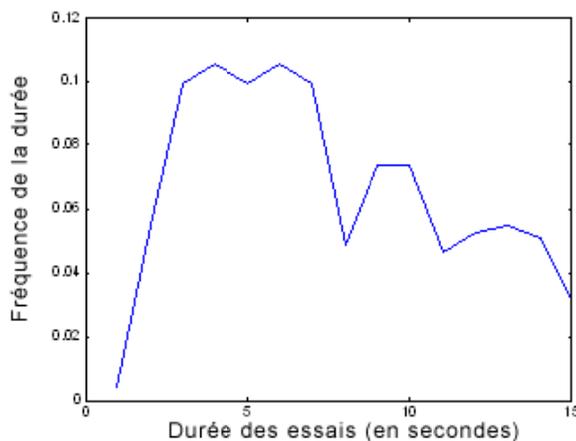


Figure 51 Distribution des durées d'essais

On voit sur la Figure 51 que la distribution des durées d'essais est centrée autour de 5 secondes. L'écart type est de 2,63s.

9.2.2 Manifestation oculomotrice de la prise de décision

9.2.2.1 Taux d'impacts sur le produit choisi

Afin de vérifier si le choix d'un produit se corrèle à l'orientation du regard, on observe l'évolution du taux d'impacts du regard sur le produit choisi en fonction du temps dans les 3 secondes qui précèdent la prise de décision. Pour chaque consigne, pour toutes les paires observées, on modélise par une sigmoïde à 4 paramètres le taux de regard sur la figure choisie.

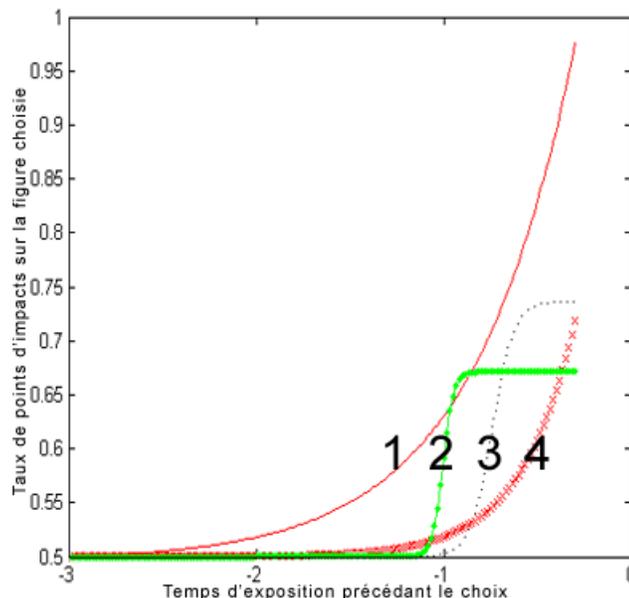


Figure 52 Modélisation des taux d'impacts sur la figure choisie avant la décision

Courbe 1 (traits pleins rouges) : Consigne « habitacle préféré » (C1+)

Courbe 2 (pointillés verts) : Consigne « habitacle le moins sportif » (C2-)

Courbe 3 (tirets noirs) : Consigne « habitacle le moins apprécié » (C1-)

Courbe 4 (croix rouges) : Consigne « habitacle le plus sportif » (C2+)

Ligne horizontale pointillée inférieure : Taux maximum d'un processus aléatoire (loi binomiale)

L'observation des courbes sigmoïdes des taux d'impacts sur le produit choisi (Figure 52) montre que dans la dernière seconde, les points de regard deviennent significativement orientés vers la modalité visuelle choisie quelle que soit la consigne, et que ce taux sera croissant ou stable. Le taux d'observation du produit choisi est maximum au moment de la prise de décision. L'hypothèse postulant l'existence d'un lien entre l'orientation du regard et le processus de décision est donc avérée, et cela quelle que soit la consigne (de préférence ou non). La consigne de préférence produit une orientation du regard en direction du produit choisi sous la forme d'un « gaze cascading effect », confirmant l'existence d'un biais potentiel dans le choix dû à une boucle de rétroaction typique à la préférence. Une étude complémentaire présentée ci-après confirme ce résultat

9.2.2.2 Répartition globale du temps d'observation sur chaque figure

Si l'on considère qu'il existe une manifestation spécifique de la prise de décision sur le regard, alors les deux figures peuvent bénéficier du même temps d'observation avant l'activation de ce processus. Le traitement précédent a montré qu'il existait bien un processus d'orientation dans la seconde précédant le choix. En comparant, pour chaque condition expérimentale, le nombre total de fixations sur la figure choisie par rapport à la probabilité (loi binomiale) de fixer une figure au hasard, aucun des deux produits n'est significativement plus observé que l'autre. Ainsi, le taux d'observation moyen du produit choisi est de 0,53 contre 0,47 pour le produit rejeté, alors que le taux devrait être supérieur à 0,57 pour être significatif. Cela confirme l'existence d'une phase d'orientation propre à la décision. La phase d'observation est donc une exploration symétrique de la scène et se concrétise par la prise de décision, accompagnée par une orientation spécifique du regard.

9.2.2.3 Durée des fixations en fonction du temps

L'étude de l'évolution des durées de fixation en fonction du temps peut permettre de distinguer la phase d'exploration et la phase de décision.

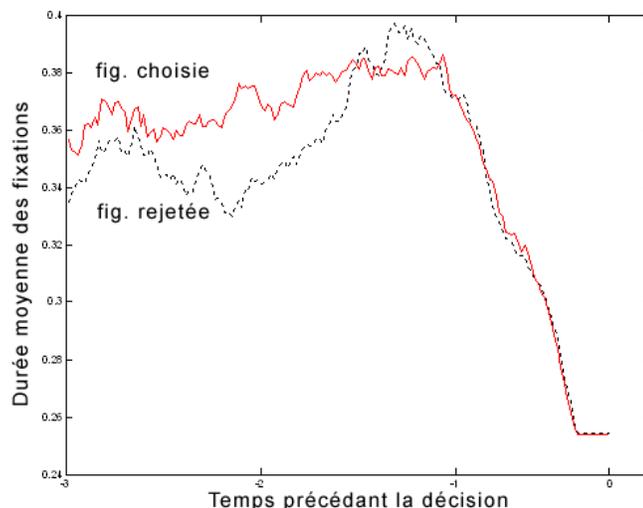


Figure 53 Evolution des durées moyennes des fixations en fonction du temps : La courbe rouge (traits pleins) représente la tendance pour la figure choisie et la courbe noire (pointillés) représente la tendance pour la figure rejetée.

On observe sur la Figure 53 qu'une seconde avant le choix la durée moyenne des fixations diminue de 300ms, alors que cette durée moyenne est stable dans les périodes précédentes, mettant ainsi en évidence un effet spécifique de la décision sur la durée des fixations. Nous n'observons cependant pas l'effet décrit dans la littérature proposant que dans la dernière seconde, la durée des fixations augmente légèrement sur la figure choisie et diminue très fortement sur la figure rejetée.

9.2.3 Spécificités de la consigne « préférence »

L'observation des formes des courbes sigmoïdes de la figure 6-2 fait ressortir que la courbe de la consigne « préférence » se distingue des autres par son taux maximum très élevée (supérieure à 95% contre 68% en moyenne pour les autres consignes), par la récence de sa croissance (la courbe croit plus tôt dans la séquence), et par la présence d'un unique point d'inflexion (caractéristique partagée avec la consigne sport+). Les courbes liées aux consignes négatives (produit le moins apprécié, et produit le moins sportif) se caractérisent par l'existence d'un deuxième point d'inflexion produisant un plateau en fin d'essai. Ces caractéristiques correspondent aux observations issues de la littérature, à savoir que la consigne de préférence produit un « gaze cascading effect » dont l'amplitude est très forte et débute tôt avant la décision. Les trois autres courbes correspondent à des choix reposant sur des critères considérés comme plus objectifs, car leur amplitude est significativement au dessus du hasard une seconde avant le choix sans être très élevée. En effet, il existe un plateau avant le choix, indiquant une stabilisation de l'exploration avant la prise de décision.

Afin de confirmer l'hypothèse que la consigne « préférence » est spécifique et que les consignes « négatives » sont plutôt similaires, on va chercher à quantifier ces similarités entre fonctions de distribution des points d'impacts sur la figure choisie au cours du temps. La distance de Kullback-Leibler est utilisée pour comparer, pour les quatre consignes, les fonctions de distribution du taux d'impacts sur la figure choisie au cours du temps. Cette distance est d'autant plus grande que les taux

évoluent différemment. Une Classification Ascendante Hiérarchique de cette distance fournit le dendrogramme suivant (Figure 54, critère de Wards sur distance euclidienne).

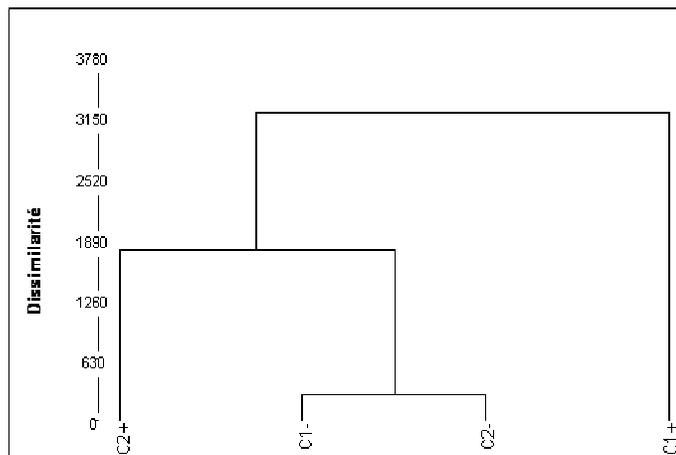


Figure 54 Dendrogramme de la Classification Ascendante Hiérarchique des distances de Kullback-Leibler entre fonctions de distributions.

On constate sur la Figure 54 que la consigne de préférence (C1+) entraîne un comportement du regard différent des trois autres consignes, car ce cas n'est pas groupé. On observe ensuite que les deux consignes négatives sont regroupées. On confirme donc que la similarité des fonctions de distribution (donc de l'orientation du regard au cours du temps) est influencée par la polarité de la consigne (rejet ou acceptation) et que la consigne de préférence est distincte des 3 autres cas. Cela confirme ainsi les observations descriptives précédentes.

Après avoir observé les différences d'orientation du regard entre consignes, nous allons observer s'il existe une prise d'information spécifique sur la figure choisie.

9.2.4 Exploration visuelle de la figure choisie

9.2.4.1 Analyse du taux de fixation de la figure choisie dans la première seconde

Pour la consigne d'évaluation du produit le plus sportif, le taux de fixation sur la figure choisie dans la première seconde est de 0,75. Pour la consigne du produit le moins sportif, ce taux est de 0,68. Pour la consigne d'évaluation du produit préféré, le taux est de 0,53 et de 0,45 pour l'évaluation du produit le moins apprécié. Avec une valeur de loi binomiale à 0,57, on observe que pour les cas d'évaluation de la sportivité, le regard s'oriente significativement vers le produit qui sera choisi pendant la première seconde de l'exposition. De plus, on constate que cela est vrai tant pour la préférence que le rejet. Cela permet de poser l'hypothèse que c'est un processus de pré-décision qui oriente le regard dans la première seconde.

9.2.4.2 Surface découverte par image

Afin d'observer si la modalité visuelle choisie est explorée plus globalement ou plus complètement que la modalité rejetée, on calcule la surface dégagée fovéalement par le regard sur chacune des deux modalités visuelles. Une analyse de variance du facteur choix (figure choisie -1-, figure rejetée -2-) et de la surface dégagée par figure, pour tous les essais, ne montre pas d'effet du facteur choix sur la valeur de cette surface ($p=0,67$). Ce traitement confirme que la paire est explorée très symétriquement ou très centralement, et que le choix ne se manifeste que sous la forme d'une évolution temporelle globale et non spatialement locale.

9.2.5 Homogénéité oculomotrice inter-essais

9.2.5.1 Similarités spatiales des distributions de fixations entre essais

Ce traitement a pour objectif de vérifier si l'exploration visuelle des zones d'intérêts est reliée aux informations requises par la consigne. Pour ce faire, la distance de Mannan va être établie entre tous les essais afin d'observer si la similarité des distributions spatiales des fixations est plus grande entre essais de la même consigne, entre essais sur la même paire, ou entre essais sur la même paire et la même consigne).

3 groupes d'essais sont construits :

- groupe 1 : ensemble des distances (de Mannan) mesurées entre tous les essais ayant été effectués sur la même paire et la même consigne
- groupe 2 : ensemble des distances mesurées entre tous les essais effectués sur la même paire, sans distinction des consignes
- groupe 3 : ensemble des distances mesurées entre tous les essais effectués sur une même consigne quelle que soit la paire

Une analyse de variance est appliquée sur le facteur « type de groupe » et la variable dépendante « distance de Mannan ». Le résultat montre ($p=0,46$) que la similarité spatiale entre essais n'est pas plus grande pour l'un de ces groupes. L'orientation locale ne semble donc pas être spécifiquement guidée (pour ce type de paradigme) par la nature de l'image ou le type de consigne.

9.2.5.2 Similarité de l'ordre des zones (géométriques) fixées entre essais

L'image est décomposée en zones géométriques sous forme d'un quadrillage. Chaque maille se voit attribué un code, et chaque fixation positionnée sur une maille reçoit le code associé. Chaque essai devient ainsi une suite de codes. La distance de Levenstein est un indice de similarité de chaîne en termes d'ordre, de présence et d'absence de zones. Les groupes définis dans la section précédente sont utilisés pour comparer la distance de Levenstein entre essais. L'analyse de variance du facteur groupe sur la variable distance de Levenstein ne montre pas d'effets significatifs. Des effets potentiels des conditions expérimentales sur la séquence de traitement visuel spatiotemporel ne sont donc pas mis en évidence.

9.2.5.3 Durées de fixations et des essais selon les participants, les images et les consignes

✓ *Durée d'essai et nombre total de fixations de l'essai*

Un test de corrélation montre que les deux variables « nombre de fixations de l'essai » et « durée d'essai » sont liées ($R=0,9$). Une analyse de variance des facteurs « image », « consigne » et « sujet » sur la variable dépendante « durée d'essai » montre que seul le facteur « sujet » ($p<0,05$) est significatif et non les facteurs « image » et « consigne ». Cela signifie qu'il y a une forte homogénéité exploratoire entre les différents essais d'un même participant. Sur plusieurs essais, un individu va explorer la scène d'une manière qui lui sera propre (idiosyncrasique) quelle que soit la paire. Notons cependant une limite : les différents essais sont effectués à la suite. Il serait pertinent de répéter ce type de tâche après plusieurs jours pour observer si cet effet idiosyncrasique persiste.

✓ *Durée de chaque fixation*

On cherche à dégager des effets des facteurs image, consigne, et participant sur la prise d'information visuelle. On s'intéresse ici aux durées des fixations. Une analyse de variance des facteurs image, consigne, et participant, sur la variable « durée de fixation » montre un effet significatif ($p<0,05$) du facteur « image » et du facteur « participant » mais pas du facteur

« consigne ». Cela signifie qu'un participant aura une tendance à explorer l'image soit avec de courtes fixations soit avec de longues fixations, et que la modalité visuelle influera sur la durée des fixations.

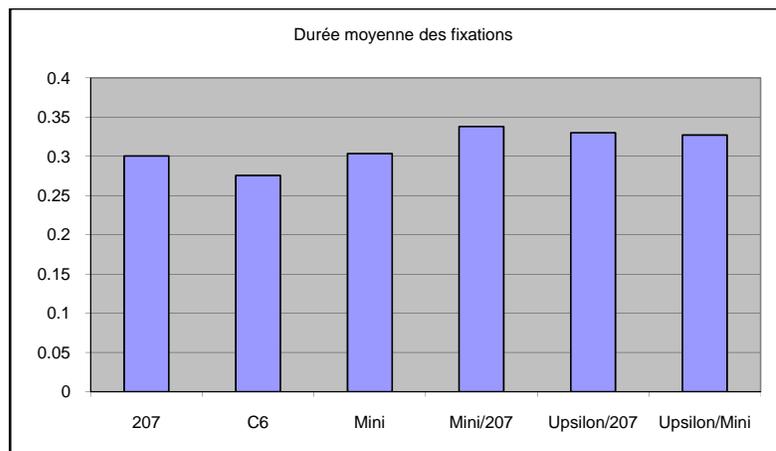


Figure 55 Moyenne des durées des fixations selon l'image

On observe que la Figure 55 que la paire influence la durée de chaque fixation, sans que la durée totale de l'essai n'en soit affectée (paragraphe précédent). Il y aurait donc une modulation de la stratégie idiosyncrasique par la nature de l'image. On note que lorsque la paire est composée de produits différents, la durée moyenne des fixations est plus grande que lorsque les paires sont composées par deux versions du même produit (mini/207, Upsilon/207, Upsilon/Mini vs. 207, C6, Mini). On peut proposer l'hypothèse que lorsque les deux éléments de la paire sont issus de produits différents, la comparaison cognitive est plus complexe. Avant de suivre un processus de comparaison, le regard serait guidé par la nécessité d'intégrer cognitivement suffisamment profondément les attributs visuels et spécificités de chaque produit. L'activité dynamique du regard serait ainsi ralentie.

✓ *Durée des fixations en fonction des zones de la scène visuelle*

Afin d'observer si la durée des fixations est liée à un processus de prise d'information locale, chaque fixation est associée à une zone géométrique d'une grille superposée à l'image. Si, pour une même image et une même consigne, certaines informations locales sont spécifiquement traitées, certaines zones peuvent être associées à des durées de fixations particulièrement courtes ou longues, où à un nombre d'entrée élevé. Le facteur zone aurait donc un effet sur la variable dépendante « durée des fixations ». Or le résultat de l'analyse de variance de la durée des fixations selon le facteur zone ne montre pas d'effets, quelle que soit la condition expérimentale.

✓ *Durée des fixations par cluster*

Un deuxième traitement est effectué en regroupant les fixations par la méthode des « Meanshift ». Pour chaque paire et chaque consigne sont constitués les groupes de fixations. On analyse ensuite la valeur des durées des fixations par groupes. Si certains clusters sont associés à un processus de prise d'information spécifique, alors il peut y avoir un effet du cluster d'attribution de la fixation sur sa durée. Une analyse de variance est appliquée sur le facteur « cluster » et la variable « durée de fixation » pour chaque condition « paire x consigne ». Il n'apparaît pas d'effets significatifs de la zone observée sur la durée des fixations. On ne met donc pas en évidence de processus de prise d'information locale caractérisés par la durée des fixations.

9.2.6 Effet oculomoteur « idiosyncrasie sujet »

9.2.6.1 Durée des essais

On a pu voir précédemment que le facteur « participant » influence significativement la durée des fixations. Cela signifie que pour tous les essais, quelle que soit la paire ou la consigne, les temps de réponse d'un participant seront significativement similaires (significativité du facteur « participant » sur la variable « durée des essais » $p < 0.05$). Un test de corrélation montre de plus que « durée d'essai » et « nombre de fixations » sont liés ($R = 0.9$). Cela signifie donc que dans une séquence d'un participant, toutes les paires recevront un nombre moyen de fixations équivalent.

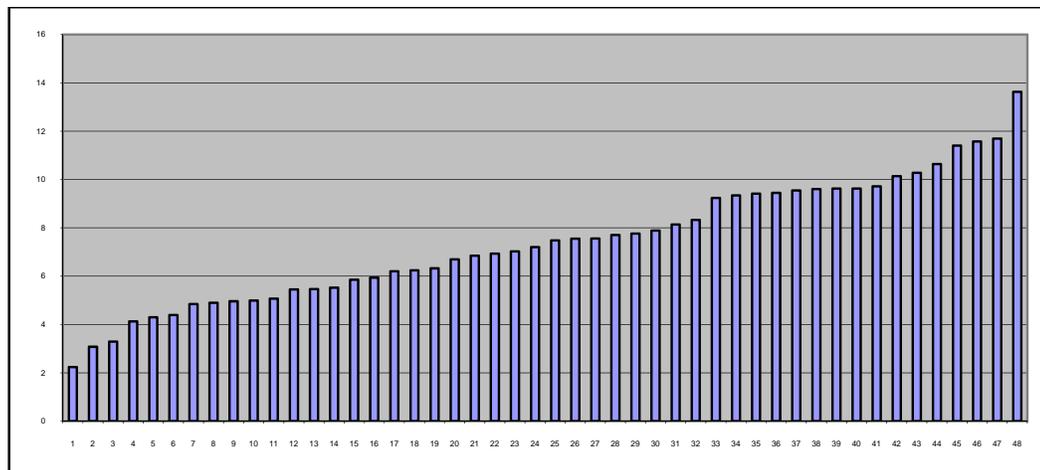


Figure 56 Moyenne des délais de réponse par participant triés par ordre croissant.

En abscisses : participants ; en ordonnées : moyenne du délai de réponse du participant

La Figure 56 montre qu'il n'existe pas de groupes de participants selon la durée moyenne des essais. On ne voit pas apparaître de plateaux ou de forts points d'inflexion. Il a été noté que les essais étaient très similaires pour un même individu en termes de répartition spatiale (distance de Mannan), de durée des fixations et nombre de fixations par essai, et de durée d'essais. Pourtant il n'apparaît pas de groupes. On ne peut donc pas considérer qu'il existe des stratégies exploratoires caractérisée par les variables précédemment citées, partagées par plusieurs individus et qui seraient liées à un mode de résolution de la consigne. Chaque individu aurait donc sa propre manière d'explorer la scène visuelle pour sur tous ses essais.

9.2.6.2 Surface dégagée

Afin de vérifier si l'idiosyncrasie individuelle se retrouve aussi pour des caractéristiques spatiales de l'exploration, la surface de l'image dégagée par voie fovéale est établie. Une analyse de variance de la surface dégagée selon les facteurs « participant », « image » et « consigne » montre que la valeur de surface dégagée est dépendante exclusivement du facteur « participant » ($p < 0,05$), mais pas de l'image ou de la consigne. Un test de Duncan ne permet pas de mettre en évidence de groupes de participants bien discriminés, les valeurs étant positionnées sur un continuum.

9.2.6.3 Durées des fixations

Une analyse de variance de la durée des fixations montre un effet significatif du facteur participant et du facteur image ($p < 0.05$) mais pas du facteur consigne. Ainsi, certains participants traitent les images avec des durées de fixations plus longues que d'autres, quelle que soit l'image, et certaines images entraînent des fixations plus longues que d'autres.

9.2.6.4 Similarité spatiale entre tous les essais d'un même participant : distance de Mannan

Afin de comparer les distances spatiales moyennes entre les essais, plusieurs groupes sont constitués : essais effectués sur une même image, essais effectués sur une même consigne, essais effectués sur une même image et une même consigne, et essais effectués par un même participant. La distance de Mannan entre chaque essai de la même classe est calculée. L'analyse de variance montre alors un effet significatif du facteur groupe sur la distance de Mannan ($p < 0.05$). Le test de Duncan montre 2 groupes de valeurs, l'un constitué des distances entre essais du groupe « même participant », le second pour les autres valeurs. On observe alors la distance moyenne de Mannan entre groupes (Figure 57).

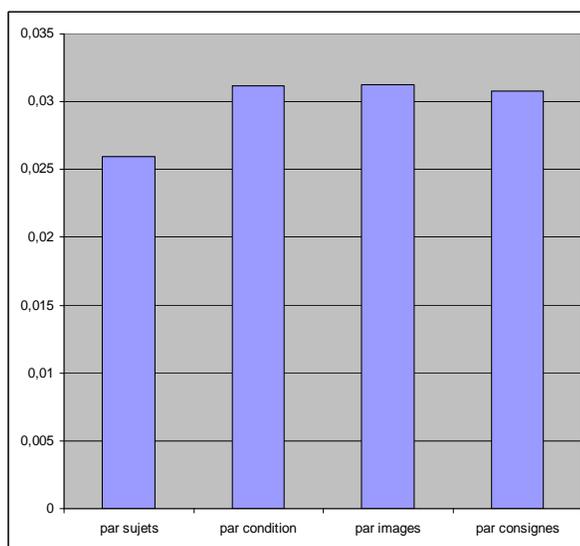


Figure 57 Distance de Mannan selon le type de regroupement des essais

Puisque la distance de Mannan est significativement plus faible entre essais d'un même participant qu'entre les essais d'un groupe « même consigne x même image », « même image » ou « même consigne », alors l'exploration visuelle est fortement guidée par des processus idiosyncrasiques. Un participant utiliserait une stratégie qui lui serait propre sur l'ensemble des essais, sans que cette dernière soit associée à la sélection d'informations visuelles car la distance entre deux essais effectués sur des paires différentes est faible.

9.2.7 Etude complémentaire

Les résultats précédents tendent à montrer que pour la consigne de préférence, le produit préféré est de plus en plus observé au cours de l'exposition. Une explication est liée au processus du « mere exposure effect » et propose que plus un objet est observé plus il est apprécié. Dans ce cas, le choix se porterait sur l'objet le plus observé. Une autre hypothèse est liée à l'attitude, où le regard s'oriente vers l'objet que l'on aime le plus. La courbe représentant le taux d'impact sur la figure choisie semble ainsi posséder les caractéristiques d'un effet cascade, manifestation d'un biais de décision par boucle de rétroaction. Afin de vérifier l'hypothèse de l'existence de ce biais, une étude complémentaire est menée en produisant artificiellement une différence de durée d'observation entre les deux produits de la paire. En reprenant le matériel expérimental de l'étude précédente, un paradigme est élaboré en créant des séquences dans lesquelles l'image de gauche est affichée alternativement avec l'image de droite. Pour une même paire, le premier élément de la paire est affiché pendant 300ms sur la partie

gauche de la scène puis le second élément est affiché sur la partie droite de la scène pendant 900ms. Cette alternance est répétée 6 fois. Une séquence similaire est construite en inversant les durées d'exposition (900ms à gauche, 300ms à droite).

Nous souhaitons ainsi vérifier si l'orientation du regard sous la forme de « gaze cascading effect » est uniquement attribuable à l'attitude ou s'il y a bien une manifestation du « mere exposure effect », et enfin si la consigne de préférence est effectivement différente de la consigne de sportivité.

Deux consignes sont donc testées : choix de l'habitacle qui est le plus apprécié, choix de l'habitacle le plus sportif.

80 essais sont effectués par consigne, à partir desquels est établi le nombre de fois que la modalité visuelle la plus longtemps affichée ait été choisie. Pour la consigne de préférence positive, la modalité visuelle la plus longtemps affichée est choisie pour 70% des essais, tandis que pour la consigne de sportivité, l'image la plus longtemps affichée est choisie à 52%. Un test de χ^2 montre alors que le produit le plus longtemps affiché est significativement préféré à l'autre ($p=0.0003$), mais qu'il n'y a pas de lien entre la durée d'exposition du produit et l'évaluation de son caractère sportif. Il y a donc bien la manifestation du « mere exposure effect » biaisant la décision : plus un produit est observé, plus il est apprécié, mais une tâche d'évaluation plus objective ne semble pas être impactée par ce biais. Il est intéressant de noter que la majorité des participants (70%) ne perçoivent pas que les deux modalités visuelles de la paire ne sont exposées sur des durées identiques (malgré un rapport de 3).

Cette étude permet donc d'étayer l'existence de deux processus distincts liés dans la prise de décision : un effet cascade du à l'observation du produit qui sera choisi, quelle que soit la consigne, et représentatif du processus de décision, mais aussi un comportement d'orientation du regard spécifique à la préférence et biaisant la décision par un jugement sur favorable au produit le plus longtemps exposé.

9.3 Synthèse

Cette étude confirme la proposition que le regard puisse être un pointeur comportemental de l'activité cognitive. En effet, il existe bien une phase de prise de décision manifestée par le regard dans les instants précédant le choix puisque la durée moyenne d'observation et la surface spatiale découverte n'est pas différentes sur chaque modalité et du fait de l'augmentation significative du taux d'impacts sur l'image choisie dans la dernière seconde, ainsi que de la diminution de la durée des fixations dans la seconde précédant le choix, se confirme l'existence d'une phase comportementale particulière dans la seconde précédant la décision. Environ une seconde avant la fin de l'essai, le taux d'impact sur la figure choisie dépasse le taux qui pourrait être attribué au hasard, et ce quelle que soit la consigne ou les images. Nous observons de plus que la consigne de préférence positive est différente des autres consignes (plus objectives ?). Elle est spécifique par la valeur maximale qu'elle atteint en termes de taux d'impacts sur la figure choisie, par l'existence d'un effet de récence (commence plus tôt), et par la présence d'un seul point d'inflexion. Il y aurait donc bien, tel que le propose la littérature, l'existence d'un biais de décision par la formation d'une boucle de rétroaction. L'étude complémentaire confirme que pour la consigne de préférence, le choix est influencé par la durée d'exposition de chaque élément de la paire, donc de l'aisance du traitement visuel de chaque modalité. Cet effet n'est pas confirmé pour la consigne de sportivité. Tel que cela a été précédemment montré, on observe que le rejet est plutôt similaire aux consignes se basant sur des critères objectifs, analytiques, tandis que la préférence se base sur des critères plutôt holistiques et hédoniques. Au

niveau de l'exploration spatio-temporelle, il n'apparaît pas que la consigne ou les modalités visuelles structurent l'exploration visuelle, qui se déroulerait en au moins deux phases : une première phase exploratoire consisterait à récolter des informations visuelles sous l'influence de nombreux facteurs attentionnels, informations qui seraient mises à la disposition des processus de prise de décision par lesquels apparaîtrait le rôle de pointeur comportemental du regard.

On note cependant que les modalités visuelles ont une influence sur la durée des fixations. Ainsi, lorsque la paire est composée de produits différents, les fixations apparaissent plus longues que sur des paires représentant le même habitacle dans une finition différente. En terme spatial, il ne semble pas que des zones bénéficient d'un traitement visuel spécifique à la consigne, et la similarité des distributions des points de fixations est faible. Un très fort effet sujet idiosyncrasique est observé. Plusieurs indicateurs de délai de réponse et d'organisation spatiale du nuage de fixations montrent qu'un même participant adopte une stratégie de prise d'information idiosyncrasique spatialement et temporellement, sans qu'il soit possible d'établir des classes comportementales sur la base de ces indicateurs.

Ces résultats amènent à proposer que quelle que soit la consigne ou la tâche, la première phase de l'exploration visuelle est associée à la collecte d'information, sans prédiction de l'importance des zones d'intérêts fixées. En fonction de propriétés cognitives idiosyncrasiques, mais aussi des spécificités visuelles de la scène, cette phase sera plus ou moins longue. S'ensuit une phase de prise de décision, qui, selon la nature de la consigne, sera influencée par un biais de « mere exposure ». Le regard ne serait donc corrélé qu'indirectement avec les informations visuelles locales nécessaires à la résolution de la consigne. Tout se passe comme si le système perceptif récoltait des informations visuelles en optimisant la prise d'information en fonction de la consigne ; des critères idiosyncrasiques définiraient ces lieux optimums ainsi que la quantité d'information requise pour le choix. Lorsque ce quota d'informations est atteint, le processus de prise de décision est activé, orientant alors le regard. Certains de ces résultats ne semblent pas confirmer ceux de la première étude où la figure choisie était plus observée que la figure rejetée, car la durée d'observation de chaque figure est ici indépendante du choix. La principale explication qui peut être avancée vient du fait que la durée des essais était fixe dans la première étude et libre dans la seconde. Ainsi, lorsque le délai est libre, le participant récolte une quantité d'information optimale depuis un point de vue optimisé, selon une stratégie propre, tandis que lorsque le délai est contraint, si celui-ci est trop long, le sujet explore l'image en détails, donc objet par objet, s'attardant sur des éléments sémantiques attractifs mais pas nécessairement pour la consigne. De plus, il était demandé dans la première étude de fixer le plus rapidement possible la figure choisie ; une figure était ainsi significativement sur-fixée dans les premières secondes, sans que l'on puisse être sûr que ce soit un processus de pré-décision qui ait orienté le regard. Avec les résultats de cette étude, on peut évoquer l'idée que pour la consigne de sportivité sur les habitacles, il puisse y avoir une pré-décision rapide sur des critères de textures et matériaux tel que l'expérience de vision contrainte l'a montré.

Cette expérience aura ainsi mis en évidence des interactions entre le comportement du regard, les demandes cognitives et les propriétés de visuelle du stimulus, sans confirmer le lien entre les informations visuelles traitée localement et les processus de jugement.

Chapitre 10

Evaluation des facteurs attentionnels

10.1 Présentation

Evaluer un produit consiste soit à se former une impression générale par des heuristiques, soit à analyser les attributs et à les comparer à des catégories cognitives préconstruites ou à comparer le produit à un autre produit de l'environnement. Puisque la verbalisation *a posteriori* du processus d'évaluation reflète plutôt une reconstruction nécessairement homogène du jugement que l'évaluation elle-même et que le regard se corrèle à l'attention, il semble intéressant d'observer quelles informations visuelles sont traitées pendant l'élaboration du jugement. Cependant, les expérimentations précédentes ont mises en évidence que les liens entre le comportement du regard, l'attention, et l'évaluation sont complexes. En effet, l'évaluation visuelle d'un produit active des processus de traitement des informations visuelles à différents niveaux (des propriétés physiques aux valeurs sémantiques), ainsi que des processus attentionnels antagonistes (attractivité de nature bottom-up vs. top-down). De plus, le traitement visuel d'une scène n'implique pas de fixer chaque objet pour l'intégrer, le regard se positionnant de manière à optimiser spatialement l'acquisition des informations visuelles. On va donc dans cette partie chercher à dégager statistiquement les facteurs qui guident l'orientation du regard.

La première partie de ce chapitre présente une méthode d'estimation des critères visuels de décision : la méthode des « Bubbles » [Gosselin01], consistant à extraire les informations locales sur lesquelles se basent la décision. Le résultat sera ensuite employé en tant que facteur de prédiction du positionnement du regard. La seconde partie de ce chapitre s'attarde sur le comportement du regard sans consigne et sa prédiction par la saillance visuelle. Enfin, dans la troisième partie, nous présentons une méthode d'estimation des contributions des facteurs de guidage, ainsi que les résultats obtenus avec nos paradigmes.

10.2 Paradigme des « Bubbles »

10.2.1 Présentation

La mesure du comportement du regard est exploitée pour mettre en évidence les zones qui ont attiré l'attention visuelle lors de l'évaluation, dans l'hypothèse où les zones observées seraient liées aux processus de jugement. Cependant, nous avons pu voir que de nombreux facteurs interviennent dans le guidage du regard. C'est pourquoi nous nous intéressons à une méthode d'estimation du caractère décisionnel de différentes zones locales de l'image, en tant que complément à la mesure oculométrique.

La méthode dite des « Bubbles » [Gosselin01] permet d'isoler les zones visuelles qui sont des critères de décision lors de l'évaluation. Elle fut initialement conçue pour l'extraction des zones

décisionnelles pour la reconnaissance des expressions faciales [Humphreys06]. Une scène est ainsi évaluée en ne laissant apparent qu'un faible nombre de zones visuelles (des 'bulles' de zones apparentes sur une image masquée). Cette méthode est pertinente lorsque la décision fait appel à des informations locales (évaluation diagnostic), ce qui doit aussi être le cas lorsque l'on souhaite inférer les critères de décision de la mesure de la position du regard. Cette technique met en jeu un traitement de corrélation entre la position spatiale des bulles et la décision du sujet. Elle nécessite un très grand nombre d'essais pour atteindre une significativité suffisante et cela d'autant plus qu'il y a de degrés de liberté dans la position des bulles. Dans les expérimentations faites par F. Gosselin, un nombre d'essais de plusieurs milliers est souvent conseillé pour un masquage dans le domaine spatial. De plus, il est nécessaire de fixer le nombre de bulles et leur surface. Ces choix s'effectuent par évaluation de tâche décisionnelle suivant les surfaces minimales des zones portant la décision (par exemple pour les expressions faciales : taille de la bulle voisine de la surface de l'œil sur l'image). Ensuite, le nombre de bulles va s'adapter à la performance des sujets durant la tâche pour tendre vers un seuil fixé de bonne décision. En moyenne, suivant la complexité de la scène, entre 10% et 15% de l'image est visible. Ensuite, le cumul des masques de bulles ayant produit une décision positive normalisé à l'ensemble des masques fournira la carte optimisant la position des bulles corrélée aux bonnes réponses.

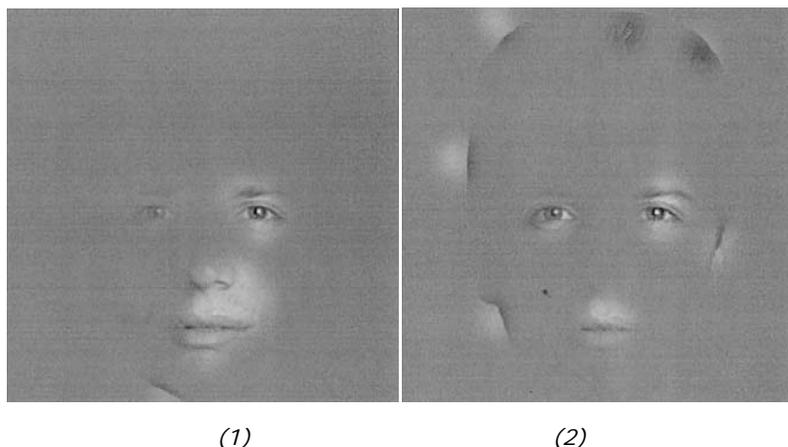


Figure 58 Images de classification des tâches de détection d'une expression et d'estimation du genre [Humphreys06]

L'image de classification (1) de la Figure 58 est issue d'une tâche de détection de la présence ou non d'une expression faciale. L'image de classification (2) est le résultat d'une tâche de détermination du genre de la personne [Humphreys06]. On observe que les zones locales de décision (zones apparentes) sont différentes selon la consigne.

Nous employons une version de comparaison par paire de cette méthode²⁵ [Humphreys06]. Plusieurs paires d'habitacles sont constituées, mais nous ne nous intéresserons qu'aux résultats sur le produit 207 et C6. La démarche est de créer plusieurs stimuli à étudier, de définir une « bonne » réponse à la tâche de discrimination, et à faire exécuter la classification par les participants. Pour le lot de paires contenant 207, la tâche est de définir quelle modalité visuelle est la plus sportive ou la plus

²⁵ L'application de cette méthode à l'évaluation monadique ne sera pas effectuée car après avoir été testée, la tâche s'est avérée soit très difficile du fait de la complexité de la consigne face au peu d'informations disponibles, soit uniquement basée sur des processus de détection/mémorisation.

haut de gamme (Figure 59), et pour le lot contenant C6, la tâche est de définir quelle modalité visuelle est la plus raffinée ou la plus haut de gamme.

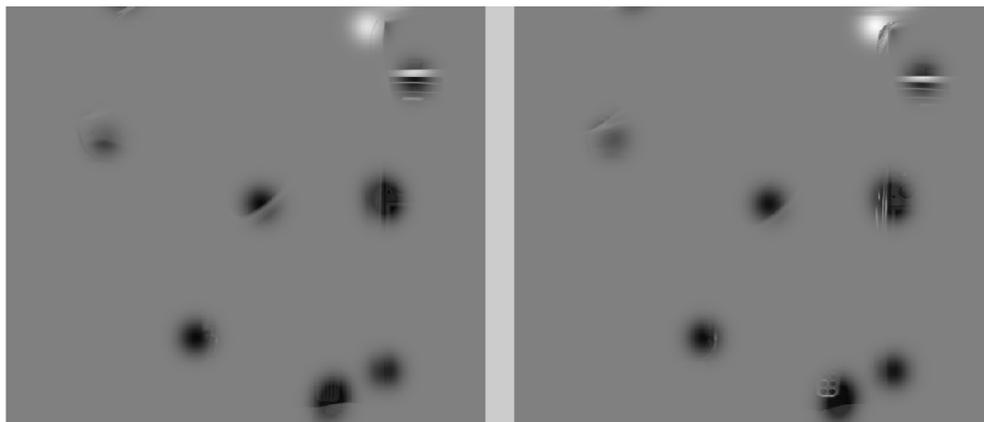


Figure 59 Exemple d'images comparées (détails) : quel est l'habitable le plus sportif ?

La tâche de classification est effectuée par un minimum de 10 participants par consigne sur au moins 250 prises de décision chacun. En combinant les réponses de tous les juges, il devient possible de définir les zones qui sont statistiquement corrélées avec les bonnes réponses.

Alors que les cartes de densité des fixations permettent d'identifier les zones attirant l'attention, le paradigme des « Bubbles » identifie les zones de décision, si la décision repose effectivement sur des zones locales et si le jugement est peu soumis à des variabilités subjectives. La méthode des « Bubbles » va faire émerger une carte de classification, c'est-à-dire une distribution spatiale des probabilités de décision avec des pics locaux bien discriminés. En comparant cette carte de classification à celle des densités de fixations, on observera s'il peut y avoir un lien entre les mécanismes attentionnelles et de jugement.

10.2.2 Résultats



Consigne sportivité

Consigne niveau de gamme

Figure 60 Images de classification

Cette méthode semble être bien adaptée au paradigme de comparaison par paire. Cependant, la qualité des résultats de classification est dépendante des conditions expérimentales. La Figure 60 représente les résultats de l'image de classification pour 207, et ces cartes semblent effectivement représenter les zones de décision : il y a donc convergence entre le choix (juste ou faux) et la position des bulles. Cependant la méthode fonctionne mal pour C6. En effet, pour la consigne niveau de gamme, le choix n'est pas homogène entre participants : l'algorithme ne peut extraire de réponses

convergentes. Pour la consigne caractère raffiné, le choix est effectivement fait en faveur de la modalité visuelle blanche, mais le seul critère est la couleur. Dès lors, la classification est uniquement un test de détection de la couleur dominante, ce qui n'est pas une information locale (Figure 61).



Figure 61 Carte de classification pour C6 consigne caractère raffiné

Pour le produit 207, les zones de décision qui se dégagent ne correspondent pas aux zones de forte densité de fixations observées dans les expérimentations 1 et 2. Les zones de décision seraient donc visualisées sans que l'attention ne s'y attarde particulièrement. La consigne de sportivité fait apparaître des zones de décision plus précises que la consigne de niveau de gamme. Cela rejoint la verbalisation post exposition des participants, qui, pour la 207 et pour la consigne de sportivité, expriment quelques zones de décision spécifique: les pédales, le levier de vitesse, les compteurs, et parfois les matériaux de la console centrale.

Ainsi, cette méthode est intéressante dans les situations suivantes :

- On peut définir une « bonne réponse »,
- Les critères de décision sont portés par des éléments visuels locaux. Si les critères sont globaux, alors ils ne seront jamais affichés dans leur totalité,
- Les critères de décision ne sont pas soumis à de fortes variabilités. S'il peut exister plusieurs bonnes réponses, c'est-à-dire si chaque élément de la paire porte des zones diagnostic antagonistes, si les critères visuels ne permettent pas de discriminer les deux modalités, ou si la décision est fortement subjective, alors l'algorithme est incapable d'extraire les zones visuelles statistiquement corrélées aux « bonnes réponses », devant recouvrir un large consensus entre les juges.

On note de plus l'apparition d'une expertise de tâche. Au fur et à mesure de la séquence, la représentation mnésique de chaque modalité visuelle s'enrichit. Lorsque cette représentation est suffisamment complète, le participant n'effectue plus un réel jugement à chaque essai, mais cherche à détecter la modalité visuelle qu'il considère comme être la bonne réponse. Avec l'apprentissage des caractéristiques visuelles de chaque modalité, le juge devient ainsi de plus en plus efficace dans la détection/catégorisation et le nombre de bulles peut ainsi devenir très faible. Une seconde limite provient de la situation où un juge inverse ses critères de décision par rapport à la « vérité terrain ». Ces essais sont alors à éliminer car bruyant les résultats. Enfin, rien n'indique que l'évaluation d'un produit sur des critères locaux (alors que les informations globales sont masquées) converge avec l'évaluation de ce même produit lorsque toutes les informations visuelles peuvent être appréhendées simultanément.

Cette méthode apporte des informations complémentaires aux enregistrements oculomoteurs en présentant les zones de décision. Cependant, il est nécessaire pour faire émerger des zones diagnostic que les critères de décision soient locaux et homogènes entre juges. Dans les autres situations, le paradigme des « Bubbles » présente peu d'intérêts. Par rapport à une simple tâche de verbalisation,

cette approche a l'avantage d'être quantitative en isolant spatialement les critères de décision par l'établissement d'une carte de classification bidimensionnelle et de dégager réellement les critères visuels de décision. Par la suite, cette image diagnostique sera utilisée comme l'un des facteurs guidant l'attention.

10.3 Exploration libre et saillance visuelle

La mesure des mouvements du regard en situation d'exploration libre consiste à placer les participants dans les mêmes conditions que les expérimentations monadiques et par paire, à la différence qu'ils reçoivent comme consigne d'observer librement la scène. Ils sont au préalable informés que l'objet affiché est un habitacle de véhicule. Ces mesures vont être utilisées en tant que référence à comparer aux situations d'évaluation avec consigne (si les distributions de fixation sont similaires avec et sans consigne, alors le facteur consigne ne dirige pas l'orientation du regard), mais aussi pour observer les effets du facteur « saillance visuelle » car en l'absence du facteur de tâche, la saillance visuelle est censée être un facteur de guidage au moins lors des premières fixations.

Une première observation du facteur « saillance visuelle » consiste à appliquer la méthode « Normalized Scanpath Saliency ». On observe alors que la prédiction de la carte de saillance est plutôt bonne lorsque l'on ne prend en compte que les 3 premières fixations de chaque séquence.

C6 Blanche	1,04
C6 Noire	1,55
207 Standard	2,36
207 Sport	3,15

Tableau 11 Valeurs NSS par habitacle

Les valeurs NSS montrent que le facteur de saillance pourrait être un bon prédicteur de l'orientation du regard en situation d'observation sans consigne et pour la première seconde d'exposition (Tableau 11). La seconde étape est de tester si les valeurs de saillance sont spécifiquement différentes sur les zones fixées par rapport aux zones non fixées sur toute la durée de l'exposition. On définit pour cela 3 ensembles de fixations : les fixations issues de la condition expérimentale, les fixations issues d'une autre condition expérimentale (image différente), et les fixations issues d'une génération aléatoire. On effectue ensuite une analyse de variance des valeurs de saillance en fonction du facteur « ensemble d'origine des fixations ». Il n'apparaît aucun effet significatif de ce facteur ($p=0,6$) sur la valeur de saillance. Ce protocole est répété pour le cas d'exploration avec consigne sans qu'il ne se dégage non plus d'effets du facteur « origine des fixations » sur les valeurs de saillance. Le facteur de saillance serait ainsi principalement actif dans la première seconde, et non quantifiable ensuite.

Alors que l'on a observé que sans consigne, le regard semble guidé par le facteur de saillance dans la première seconde d'exposition, on compare les valeurs de saillance des fixations avec consigne et des fixations sans consigne, sur toute la durée d'exposition. Un test de Student est appliqué pour comparer les deux distributions de valeurs. Selon ce test ($\alpha=0,4$), les valeurs ne proviennent pas de distributions différentes. Plusieurs interprétations sont possible : le facteur de saillance visuelle est tout autant influent ou tout aussi peu influent avec que sans consignes, ou le facteur consigne est tout autant influent ou tout aussi peu influent avec que sans consignes. D'après ces observations, les propriétés des signaux des zones de fixations ne diffèrent pas avec et sans consigne. On peut donc

proposer que le facteur de saillance ne soit pas très efficace pour prédire les zones de fixations pour ce type de paradigmes sur les durées d'exposition employées.

10.4 Evaluation des facteurs attentionnels suivant un modèle de mélange additif

10.4.1 Présentation

L'objectif de cette section est de présenter un algorithme d'estimation de la contribution de différents facteurs de guidage du regard sur la distribution spatiale des fixations oculaires. Cette méthode, proposée par Vincent, Correani, Baddeley, Troscianko, et Leonards (Bristol University [Vincent07]) consiste à considérer que la densité des fixations peut s'écrire comme un mélange additif de densités élémentaires, chacune représentant un facteur de prédiction des positions du regard.

Pour cela, un algorithme de type « Expectation-Maximisation » est utilisé. Le mélange comprendra les fonctions de densité choisies *a priori*. A ce mélange, des fonctions de densité gaussiennes pourront être ajoutées suivant les hypothèses à vérifier. A la convergence, on recueille la contribution de chaque densité dans le mélange et la paramétrisation des modes gaussiens (moyenne, matrice de variance-covariance).

Pour définir correctement le modèle de mélange, il est nécessaire de mettre en adéquation les hypothèses choisies avec les images utilisées. Ainsi dans son article, Vincent choisit les hypothèses suivantes sur des images de scènes urbaines nocturnes : distribution uniforme, attractivité du premier plan, biais d'observation centrale de la scène, attractivité des zones lumineuses, attractivité des contours de zones lumineuses, attractivité du ciel, attractivités des zones de forts contrastes.

Dans le cadre de nos paradigmes, nous choisissons les facteurs de guidance suivants :

- Hypothèse de distribution uniforme (aléatoire) : chaque zone de l'espace possède la même probabilité de fixation. Elle est définie sous la forme d'une distribution uniforme captant des fixations sur la totalité de la surface de l'image. Dans le mélange, cette carte joue le rôle de carte « par défaut » captant les fixations qui ne sont pas expliquées par les autres hypothèses.

- Hypothèse de centralité : l'observation d'images sur écran produit un biais de centralité. Cela signifie que le regard se pose préférentiellement au centre de l'écran et a tendance à y revenir régulièrement, indépendamment du contenu de l'image ou de l'activité cognitive. Une carte de centralité est donc définie, où les zones centrales ont une probabilité de fixation plus élevée que les zones périphériques. La carte de densité du facteur centralité est établie par une fonction gaussienne appliquée au centre de l'image, la valeur diminuant lorsque l'excentricité augmente. Les paramètres de ce mode sont fixés et restent constants (Figure 62).



Figure 62 Exemple du modèle de centralité sur le cas monadique 207-sport

- Hypothèse de saillance visuelle : En modélisant les propriétés de traitements du signal du système visuel, il est possible de prédire l'attractivité relative des différentes zones du champ visuel en l'absence de facteurs attentionnels « top-down ». On va ainsi établir une carte de saillance, définissant les zones vers lesquelles le regard s'orientera préférentiellement. L'un des principes de base est que le regard est attiré par les zones de forte variance texturale (fréquence, orientation), de luminance et chrominance. Cette carte est établie par le logiciel de saillance développé au Laboratoire des Images et Signaux de l'INP Grenoble (Figure 63) – présentation en annexe.



Figure 63 Illustration de la carte de saillance : image 207-sport

- Hypothèse de maximisation de l'information (Figure 64) : ce facteur de prédiction du positionnement du regard provient de l'observation expérimentale [Renninger05b] que le regard ne se pose pas nécessairement sur une zone alors que l'attention la traite, mais qu'il se positionne de telle sorte à optimiser spatialement l'acquisition d'informations visuelles. Cette approche se base sur le concept de centre de gravité informationnel. La carte de maximisation se construit par l'extraction des contours des objets de scène via un filtrage de Sobel, puis par regroupement de ces contours par la méthode des K-Mean. On obtient alors des barycentres de contours, depuis lesquels le regard est à équidistance de plusieurs objets.

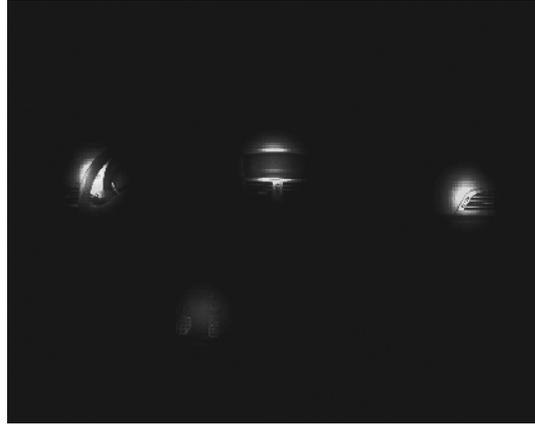


Figure 64 Illustration de la carte de maximisation pour l'image 207-sport

- Hypothèse sémantique : à l'opposé de l'hypothèse de guidage du regard par un facteur de saillance, on peut proposer que les mouvements du regard soient prédits par le contenu sémantique de la scène. Puisque l'on est ici en situation d'évaluation, la carte de pertinence sémantique va se construire sur la base de la carte de classification issue du paradigme des « Bubbles ». Pour le produit « C6 consigne niveau de gamme » pour lequel il n'y a pas de carte de classification, la carte sémantique est associée à la carte de classification de la consigne « caractère raffiné » car il s'est avéré dans les verbalisations que les deux consignes faisaient appel aux couleurs.



Figure 65 Illustration de la carte sémantique pour l'image 207-sport

Ces 5 hypothèses de guidage du regard fournissent ainsi 5 cartes de densité prédictives (de probabilité spatiale de fixation), dont le mélange additif expliquera la densité mesurée. On considère que ces 5 cartes sont indépendantes entre elles. Elles doivent de plus avoir des distributions suffisamment différentes pour ne pas avoir de valeurs similaires en un même point. Ces cartes de prédiction sont construites pour chaque condition expérimentale des expérimentations monadique et par paire. La première étude porte sur les cas d'exploration libre des images étudiées dans les expérimentations 1 et 2 (sans consigne). La seconde étude porte sur les conditions expérimentales (images et consignes) des expérimentations 1 et 2.

10.4.2 Résultats sur les distributions des fixations issues de l'exploration libre

Dans cette première partie nous observons les contributions de chaque hypothèse de guidage du regard à la distribution spatiale des fixations mesurées en situation d'exploration libre (sans consigne).

✓ Evaluation des contributions pour le paradigme d'exposition monadique

Le Tableau 12 présente le taux de contribution de chacune des hypothèses de prédiction par image.

condition/facteur	centralité	maximisation	saillance	sémantique	uniforme
<i>207-Sport</i>	0,62	0,21	0,00	0,18	0,00
<i>207-Standard</i>	0,45	0,23	0,00	0,31	0,00
<i>C6-Noire</i>	0,67	0,22	0,04	0,08	0,00
<i>C6-Blanche</i>	0,77	0,19	0,00	0,04	0,00

Tableau 12 Estimation des contributions des différents facteurs de guidage pour l'exploration libre

On observe tout d'abord que le facteur centralité est très explicatif pour les 4 cas, c'est-à-dire que la contribution de ce facteur est élevée (0.62, 0.45, 0.67, 0.77). Pour les produits 207, le facteur sémantique est influent (0.17, 0.31). Or la carte sémantique est construite sur la base de la carte de classification des Bubbles, donc des zones de décision. Cela signifie que s'il existe une consigne ou non, les zones sémantiques de décision attirent le regard. On constate que la carte de saillance n'a quasiment pas d'effets, mais que la carte de maximisation en a. Enfin, la contribution du mode uniforme est nulle, indiquant que le regard ne se positionne pas aléatoirement sur l'image. Afin d'analyser plus en détail les effets des différents facteurs, le calcul est effectué de nouveau en éliminant le facteur centralité qui capte entre la moitié et les trois-quarts des fixations.

Le Tableau 13 présente les résultats de l'évaluation de la contribution des différents facteurs de guidage sans le facteur centralité.

condition/facteur	maximisation	saillance	sémantique	uniforme
<i>207-Sport</i>	0,24	0,05	0,71	0,00
<i>207-Standard</i>	0,30	0,00	0,70	0,00
<i>C6-Noire</i>	0,40	0,46	0,15	0,00
<i>C6-Blanche</i>	0,39	0,58	0,03	0,00

Tableau 13 Taux de contribution des facteurs sans l'hypothèse de centralité

Alors que l'on aurait pu penser que les facteurs de maximisation et de saillance soient fortement contributeurs de la distribution des fixations dans une situation d'exploration libre (sans consigne évaluative), il apparaît pour le produit 207 que c'est la carte sémantique qui explique le plus l'orientation du regard. Pour le produit C6, c'est la saillance et la maximisation. Ainsi, les zones qui sont sémantiquement attractives avec consigne le sont aussi sans consigne, et en l'absence de telles zones (cas C6), les facteurs de saillance et de maximisation sont explicatifs de la distribution des fixations.

Si les fixations sur les produits 207 sont plutôt expliquées par la carte sémantique, la carte de saillance intervient pour le cas 207-sport et pas pour la condition 207-standard. Ainsi, lorsque la modalité visuelle est perceptivement plus complexe à traiter (C6-noire) ou qu'il n'y a que peu de zones sémantiques attractives (207-standard), alors le facteur de maximisation est plus fort. Dans le cas de zones visuelles sémantique attractives (associées à une forte saillances ?) et dans le cas d'une modalité visuelle plus aisément traitable (207-sport et C6-blanche), le facteur de saillance contribue plus que dans le cas précédent.

Nous allons maintenant observer les différentes contributions pour les situations d'exposition par paire.

✓ *Evaluation des contributions pour le paradigme d'exposition par paire*

L'évaluation des contributions des facteurs de guidage du regard lors de l'exploration libre des paires fournit le tableau de synthèse suivant.

condition/facteur	centralité	maximisation	saillance	sémantique	uniforme
<i>207-Paire1</i>	0,08	0,08	0,00	0,84	0,00
<i>207-Paire2</i>	0,16	0,05	0,00	0,79	0,00
<i>C6-Paire1</i>	0,00	0,32	0,54	0,13	0,00
<i>C6-Paire2</i>	0,15	0,52	0,24	0,09	0,00

Tableau 14 Taux de contribution des 5 facteurs

On observe dans le Tableau 14 que le facteur de centralité explique peu la contribution des fixations. Cela est cohérent dans une exposition par paire. Le mode uniforme est de nouveau nul, indiquant que les distributions des fixations ne sont pas aléatoires. La carte sémantique explique bien les distributions sur les produits 207, tandis que les facteurs maximisation/saillance expliquent les distributions des produits C6. Afin d'étudier plus précisément les tendances qui émergent, le calcul des contributions est relancé en éliminant le facteur centralité.

condition/facteur	maximisation	saillance	sémantique	uniforme
<i>207-Paire1</i>	0,09	0,04	0,88	0,00
<i>207-Paire2</i>	0,09	0,03	0,88	0,00
<i>C6-Paire1</i>	0,32	0,54	0,13	0,00
<i>C6-Paire2</i>	0,55	0,37	0,09	0,00

Tableau 15 Taux de contribution des facteurs sans l'hypothèse de centralité

Nous pouvons voir dans le Tableau 15 que même sans consigne les zones sémantiques de décision pour la consigne permettent d'expliquer la distribution des fixations pour le produit 207, alors que l'on aurait pu penser que les facteurs de maximisation et de saillances auraient été les principaux guides de l'exploration de l'image. Cela signifie donc que la carte sémantique ne relève pas uniquement la pertinence des zones au regard de la consigne, mais plutôt un potentiel d'attractivité sémantique local. L'attractivité de ces zones pourrait être due à un processus de distinctivité ou d'identification du produit. Dans le cas du produit C6, on trouve que les facteurs saillance et maximisation participent tous les deux la distribution des points de fixation, et que la saillance contribue plus pour le cas Paire 1 que paire 2, où c'est la maximisation qui prime. On constate aussi que le facteur uniforme est nul pour les quatre cas.

Ainsi, lorsqu'il y a dans l'image des attracteurs attentionnels locaux, comme dans le cas de 207, ce sont ces zones qui vont guider l'exploration visuelle de l'image, qu'il y ait une consigne ou non. Dans le cas où il n'y a pas de zones locales attractives, l'exploration va être guidée par les facteurs de saillance et de maximisation.

Pour expliquer le fait que le facteur de saillance contribue plus pour C6 sur la paire 1 que sur la paire 2, on peut avancer l'idée que lorsque le produit à couleur dominante blanche est à droite celui-ci est vu sans difficultés, alors que le produit de gauche qui est noir sera difficilement perçu. Le regard se positionnera alors sur les zones les plus saillantes de la modalité de gauche car ce sont celles-ci qui seront les plus aisément détectées en périphérie. Lorsque la modalité blanche est à gauche, le produit noir est aisément traité en vision central, et la vision périphérique est suffisante pour extraire les principales informations configurales de la modalité blanche. Dans ce cas, ce ne sont pas les zones de

saillances qui seront sélectionnées, mais les zones maximisant l'extraction de l'information visuelle, sur la base d'une pré-connaissance configurale.

On retient de ces mesures que la carte sémantique est explicative des distributions de 207 lorsque l'exploration est libre. Pour C6, les facteurs saillance et maximisation sont tout deux influents. Dans aucun des cas on ne note d'effets du facteur uniforme.

Afin d'aller plus loin dans cette étude, nous allons désormais étudier les contributions des différents facteurs de guidage pour les cas d'évaluation monadique et par paire avec consigne.

10.4.3 Résultats sur les distributions des fixations issues des expérimentations avec consignes

On propose d'intégrer dans les modèles de prédiction une carte qui à la différence des autres hypothèses, indépendante entre elles, peut être dépendante de 4 des 5 facteurs listés²⁶. Ce nouveau modèle est la carte de densité des fixations mesurée lors de la tâche d'exploration libre. En effet, si la carte de densité des fixations mesurées en situation d'exploration libre explique la distribution des fixations avec consigne, on pourra considérer que les effets de la tâche sont soit subtiles, soit annulés par des divergences (non homogénéité) inter-sujets, soit indépendants du positionnement du regard. Si au contraire la carte de première impression n'explique pas la distribution avec consigne, alors on pourra considérer que le facteur « consigne » a un effet sur l'orientation de l'attention visuelle ; il sera alors intéressant de déterminer la nature de cet effet.

10.4.3.1 Evaluation des contributions en situation d'exposition monadique

✓ *Prise en compte des facteurs « centralité », « maximisation », « saillance », « exploration libre » sémantique » et « uniforme »,*

L'évaluation des facteurs attentionnels par l'algorithme fournit les résultats du Tableau 16.

condition / facteur	centralité	maxi.	saillance	explo. libre	sémantiq.	uniforme
207-sport consigne sportivité	0,15	0,09	0	0,67	0,09	0
207-sport consigne niveau de gamme	0	0,03	0	0,97	0	0
207-standard consigne sportivité	0	0,08	0	0,83	0,09	0
207-standard consigne niveau de gamme	0,4	0,13	0	0,47	0	0
C6-blanche consigne raffinement	0,31	0,07	0	0,58	0,04	0
C6-blanche consigne niveau de gamme	0,09	0,06	0	0,76	0,08	0
C6-noire consigne raffinement	0,03	0,05	0,02	0,87	0,03	0
C6-noire consigne niveau de gamme	0,32	0,05	0	0,61	0,02	0

Tableau 16 Taux de contribution des différents facteurs

Le facteur « exploration libre » apparaît comme le principal prédicteur de la densité des fixations ; il est anti-corrélé au facteur centralité ; à eux deux, ces facteurs expliquent au minimum 80% de la distribution. Le facteur « uniforme » a une contribution nulle, la distribution des fixations n'est donc pas de nature aléatoire. Ainsi, l'exploration visuelle de l'image est très bien prédite par des facteurs

²⁶ . En l'absence de consigne, le regard peut être guidé par un facteur uniforme, de biais de centralité, de saillance et de maximisation, mais aussi d'attractivité sémantique sans rapport à la consigne

indépendants de la consigne. Pour le produit 207 que le modèle sémantique explique 9% des fixations de la distribution pour la consigne de sportivité mais 0% pour l'autre consigne. Il y aurait donc ici la mise en évidence d'un effet de la nature de la consigne sur la contribution du facteur « sémantique ».

Cependant, la carte de prédiction par « exploration libre » est dépendante des autres hypothèses ; On relance donc l'évaluation des contributions en éliminant ce facteur.

✓ *Prise en compte des facteurs « centralité », « maximisation », « saillance » et « pertinence »*

condition / facteur	centralité	maximisation	saillance	sémantique	uniforme
207-sport consigne sportivité	0,41	0,15	0,00	0,44	0,00
207-sport consigne niveau de gamme	0,50	0,15	0,00	0,34	0,00
207-standard consigne sportivité	0,24	0,24	0,00	0,53	0,00
207-standard consigne niveau de gamme	0,63	0,19	0,00	0,18	0,00
C6-blanche consigne raffinement	0,61	0,19	0,12	0,08	0,00
C6-blanche consigne niveau de gamme	0,56	0,26	0,06	0,12	0,00
C6-noire consigne raffinement	0,52	0,31	0,12	0,05	0,00
C6-noire consigne niveau de gamme	0,73	0,22	0,00	0,05	0,00

Tableau 17 Taux de contribution des différents facteurs sans prise en compte de l'hypothèse d'« exploration libre »

On voit dans le Tableau 17 que lorsque la consigne oriente l'attention vers des critères visuels locaux (consigne sportivité), ou lorsque certains attributs du produit émergent particulièrement en termes de distinctivité (207-sport), la distribution des fixations s'explique par la carte de centralité et la carte sémantique, tandis que dans les autres cas, c'est la carte de centralité et de maximisation qui expliquent les distributions. Ainsi, lorsque l'attention est attirée par des zones spécifiques (soit pour des raisons de pertinence de consigne, soit de distinctivité), le regard s'oriente sur ces zones. Sinon, le regard se positionne depuis un point de vue optimal pour traiter un maximum d'informations visuelles. On peut donc distinguer deux types de situations : celles où le regard s'oriente vers des zones spécifiques, et celle où il se positionne sur des zones de maximisation des informations visuelles.

✓ *Prise en compte des facteurs maximisation, saillance, sémantique et pertinence*

Dans cette troisième et dernière étape, nous éliminons le facteur « centralité » afin de comparer les trois facteurs minoritaires : maximisation, saillance et sémantiques. Le mode uniforme est toujours utilisé pour capter le cas échéant les fixations qui seraient associées à une distribution aléatoire.

condition / facteur	maximisation	saillance	sémantique	uniforme
207-sport consigne sportivité	0,17	0,00	0,83	0,00
207-sport consigne niveau de gamme	0,16	0,06	0,78	0,00
207-standard consigne sportivité	0,26	0,00	0,74	0,00
207-standard consigne niveau de gamme	0,25	0,05	0,70	0,00
C6-blanche consigne raffinement	0,33	0,51	0,13	0,00
C6-blanche consigne niveau de gamme	0,45	0,40	0,15	0,00
C6-noire consigne raffinement	0,45	0,46	0,10	0,00
C6-noire consigne niveau de gamme	0,40	0,47	0,13	0,00

Tableau 18 Taux de contribution des facteurs maximisation, saillance, sémantique et uniforme

On observe sur le Tableau 18 que les produits C6 et 207 sont bien discriminés, principalement par le facteur sémantique. Ainsi, les distributions des fixations sur 207 s'expliquent par une forte contribution du facteur sémantique tandis que les fixations sur C6 s'expliquent par les facteurs saillance et maximisation. Cela semble cohérent car il y a peu de zones sémantiques locales pour C6 contrairement à 207. On observe ensuite que la modalité sport de 207 se discrimine de la modalité standard, car le facteur « maximisation » est plus fort dans le second cas. Cela peut se justifier par le

fait que pour le produit 207-sport, les fixations et les zones de décision correspondent mieux. Une hypothèse complémentaire est qu'en l'absence d'attributs de distinctivité, le regard se positionne sur des centres de maximisation. Pour C6, il n'apparaît pas de discrimination des produits selon la modalité visuelle ou la consigne.

✓ *Synthèse*

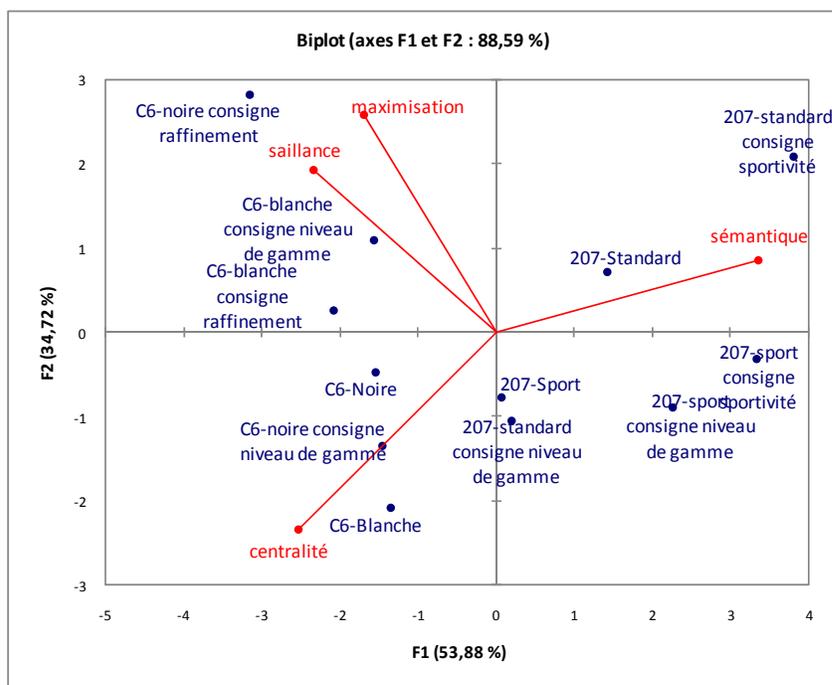


Figure 66 Analyse en Composantes Principales des contributions relatives des facteurs de guidage du regard pour la situation d'exposition monadique

Ces résultats ont montré tout d'abord qu'il existe une très forte tendance à explorer l'image depuis un point de vue central quelle que soit l'image (puissance du facteur « centralité »), que la distribution des fixations n'est pas aléatoire (absence du mode uniforme), et que la consigne semble avoir un faible effet sur l'orientation du regard (bonne prédiction du modèle de densité de l'exploration libre). Sur ce type de produits et de consignes, il est donc difficile de mettre en évidence des zones vers lesquelles s'est porté l'attention pour des motivations de jugement. Cependant, lorsqu'il existe des attributs visuels distincts (se différenciant du proto-objet mnésique) ou encore des attributs pertinents cognitivement (critères de décision), la carte sémantique a un plus fort effet que celles de maximisation et de saillance.

C'est ainsi que l'on observe sur l'Analyse en Composantes Principales des individus « situation d'exposition monadique » sur les variables « centralité, saillance, maximisation et sémantique » (Figure 66) que l'axe F1 (53% de l'inertie) est principalement expliqué par le facteur sémantique, qui discrimine les produits C6 et 207. Il discrimine aussi les conditions d'observation de 207 avec consigne sportivité des autres consignes. Ainsi, en toute cohérence, la carte de décision est mieux explicative lorsque la consigne oriente l'attention visuelle sur des signes locaux. Notons que l'axe F2 (35% de l'inertie) discrimine plutôt bien les 207-sport des 207-standard.

Cette méthode est donc efficace pour comprendre les facteurs qui guident l'attention, mais le chevauchement spatial de plusieurs facteurs doit modérer l'interprétation des effets pour les différentes conditions expérimentales. Enfin, puisque la participation de chaque facteur est très

variable selon ces conditions expérimentales, il n'est pas possible de construire un modèle de prédiction global robuste sur la base de ces données.

10.4.3.2 Evaluation des contributions en situation d'exposition par paire

✓ *Prise en compte des facteurs « centralité, maximisation, saillance, exploration libre, sémantique et uniforme »*

condition / facteur	centralité	maximisation	saillance	expl. libre	sémantiq.	uniforme
207-Paire1 consigne sportivité	0,00	0,00	0,00	0,87	0,13	0,00
207-Paire1 consigne niveau de gamme	0,05	0,00	0,00	0,46	0,49	0,00
207-Paire2 consigne sportivité	0,00	0,00	0,00	0,72	0,28	0,00
207-Paire2 consigne niveau de gamme	0,00	0,03	0,00	0,75	0,23	0,00
C6-Paire1 consigne raffinement	0,06	0,04	0,00	0,87	0,02	0,00
C6-Paire1 consigne niveau de gamme	0,00	0,08	0,00	0,92	0,00	0,00
C6-Paire2 consigne raffinement	0,00	0,30	0,00	0,70	0,00	0,00
C6-Paire2 consigne niveau de gamme	0,03	0,18	0,00	0,77	0,01	0,00

Tableau 19 Taux de contribution des facteurs de guidage du regard

On observe sur le Tableau 19, comme dans le cas de l'exposition monadique, que le facteur « exploration libre » est de loin le plus fort contributeur à l'explication de la distribution des fixations, à l'exclusion du cas « 207-paire 1, consigne niveau de gamme ». Cela signifie que les effets de la consigne sont plutôt faibles. Contrairement à l'exposition monadique, l'effet du facteur « centralité » est faible. On a effectivement pu observer précédemment que le biais de centralité n'est pas positionné au centre de la scène en situation d'exposition par paire mais au centre de chaque modalité visuelle.

Enfin, on observe que la carte sémantique est le second contributeur pour les distributions de fixations pour le produit 207, alors que c'est la carte de maximisation qui l'est pour le produit C6 ; ces résultats sont similaires au cas d'évaluation monadique.

Comme lors de l'étude de l'évaluation monadique, on réitère les évaluations en éliminant le modèle d'exploration libre car il inclut les autres facteurs, ainsi que le modèle de centralité car sa pertinence est faible dans cette situation d'exposition.

✓ *Prise en compte des facteurs « maximisation, saillance, sémantique et uniforme »*

L'évaluation des facteurs par l'algorithme fournit les résultats exposés dans le Tableau 20.

condition / facteur	maximisation	saillance	sémantique	uniforme
207-Paire1 consigne sportivité	0,03	0,00	0,97	0,00
207-Paire1 consigne niveau de gamme	0,01	0,03	0,95	0,00
207-Paire2 consigne sportivité	0,06	0,00	0,94	0,00
207-Paire2 consigne niveau de gamme	0,16	0,00	0,84	0,00
C6-Paire1 consigne raffinement	0,44	0,52	0,04	0,00
C6-Paire1 consigne niveau de gamme	0,50	0,49	0,00	0,00
C6-Paire2 consigne raffinement	0,80	0,12	0,08	0,00
C6-Paire2 consigne niveau de gamme	0,60	0,32	0,07	0,00

Tableau 20 Taux de contribution des facteurs maximisation, saillance, sémantique et uniforme

On observe dans le Tableau 20 que pour le produit 207, la carte sémantique est le meilleur modèle de prédiction. On doit noter que les zones sémantiques de décision sont aussi des zones qui diffèrent entre les deux modalités visuelles, ne permettant pas de savoir si ces zones sont observées pour leur différence ou pour leur significativité. On observe de plus que les paires 1 (modalité visuelle « sport » à gauche dans la paire) sont discriminées des paires 2 par le facteur maximisation, plus faible pour la

paire 1 que pour la paire 2. Cela pourrait s'expliquer par le fait que lorsque le regard se repositionne sur la modalité sport à gauche (paire 1), il se pose directement sur des zones de décision car celles-ci sont perceptivement aisées à extraire. Par contre lorsque c'est la modalité standard qui est à gauche, le regard se positionne depuis un point de vue optimal (maximisation).

Pour le produit C6, le facteur de maximisation domine pour la paire 2 (modalité blanche à gauche) alors que pour la paire 1, les facteurs de saillance et de maximisation sont à égalité. La carte sémantique a un effet limité car il n'y a que peu d'attributs locaux qui répondent directement aux consignes. On peut poser l'hypothèse que lorsque c'est la modalité blanche qui est en périphérie (Paires 2 ; position excentrée car poste de conduite éloigné du centre de l'image), le système visuelle acquiert facilement des informations permettant, lors du repositionnement du regard de sélectionner un point de vue optimal, d'où un fort effet de la maximisation. Lorsque c'est la modalité noire qui est à gauche (paires 1), alors il est difficile par la vision non fovéale d'analyser perceptivement le stimulus, et c'est alors la carte de saillance, de plus bas niveau de traitement que la carte de maximisation, qui oriente le regard.

Ces résultats ne mettent pas en évidence le fait que l'attention s'oriente spécifiquement vers les critères visuels de décision, même s'ils témoignent d'un effet de la pertinence sémantique sur l'attention. Enfin, ici encore on note l'absence de la contribution du mode uniforme.

✓ Synthèse

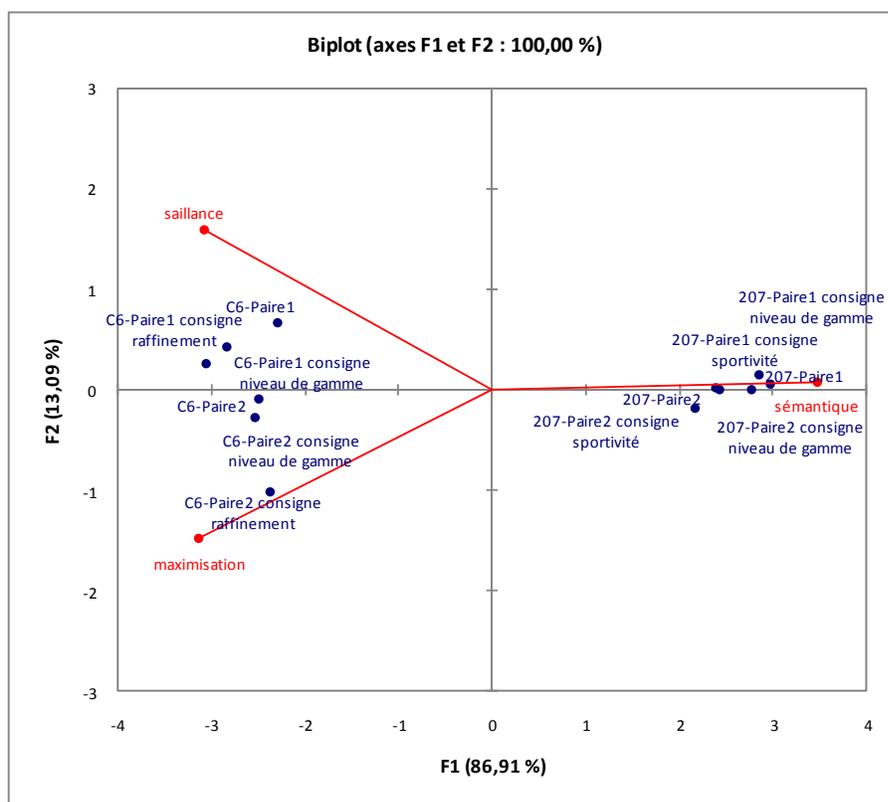


Figure 67 Analyse en Composantes Principales des contributions relatives des facteurs de guidage du regard pour la situation d'exposition par paire

La Figure 67 représente l'Analyse en Composantes Principales des différentes conditions expérimentales par paire avec et sans consigne en fonction des valeurs de contribution des facteurs « maximisation », « saillance » et « sémantique ». On remarque tout d'abord que l'axe F1 contient

presque toute l'inertie (87%). Cela indique que la principale discrimination est effectuée selon la valeur de contribution de la carte sémantique. Ainsi, les paires contenant les 207 sont plutôt associés à une forte composante sémantique et les paires contenant C6 à une faible composante sémantique. De plus, on observe que l'axe F2 (13% de l'inertie), expliqué par les facteurs saillance et maximisation, discrimine les paires 1 des paires 2 pour les deux produits. Ainsi, qu'il y ait des consignes ou non, c'est le type de produit représenté et la position relative dans la paire de ces produits qui expliquent les distributions des fixations du regard. Cela explique pourquoi on observe comme pour le cas monadique que le facteur « première impression » est le principal facteur expliquant la distribution des fixations : la scène visuelle représentée guide très fortement le regard quelle que soit la consigne.

On observe enfin dans les cas avec C6 que la paire 2 est associée à un facteur de maximisation plus élevé que la paire 1, et la saillance étant plus élevée pour cette paire 1. Ainsi, lorsqu'il y a peu d'informations locales spécifiques aux consignes tant visuellement que cognitivement que la position de la modalité visuelle la plus aisée à traitée perceptivement (mais aussi celle qui est choisie...) implique une différence dans les effets de la saillance et de la maximisation.

10.4.3.3 Conclusion

Cette méthode semble très intéressante pour relever la contribution relative de différents facteurs influençant l'orientation du regard. Sans établir de critères de décision sur la robustesse des effets, cette méthode descriptive permet de comparer les différentes conditions expérimentales pour dégager des spécificités attentionnelles. Les résultats ne semblent pas indiquer que le regard s'oriente spécifiquement sur des zones visuelles décisionnelles, mais plutôt sur des zones de distinctivité, car les prédictions des effets respectifs des facteurs suivent les mêmes tendances avec et sans consigne. Ces zones sémantiques peuvent ainsi être des critères de jugement si la consigne est orientée vers des informations visuelles locales, mais peuvent aussi être des zones attractives sur des critères d'innovation, de préférence, de complexité visuelle, ou encore d'identification. Dans l'estimation des facteurs de guidage du regard en situation de comparaison par paire, il serait pertinent de remplacer la carte de centralité globale par une carte à deux gaussiennes, chacune centrée sur l'une des modalités visuelles de la paire, car il s'agit du meilleur modèle de mélanges de gaussiennes, donc fortement représentatif de la distribution. Au niveau de l'algorithme, il semble qu'un biais d'incertitude computationnelle puisse apparaître lorsqu'en un point les valeurs des cartes de prédiction sont similaires. La pertinence d'attribuer la fixation à un seul facteur est alors limitée, faussant les effets relatifs des différents facteurs. Cette méthode, qui n'en n'est qu'à ses débuts, semble très prometteuse pour l'étude des processus attentionnels.

Chapitre 11

Synthèse de la partie expérimentale

La première partie de notre travail expérimental est constitué de deux paradigmes d'évaluation du produit automobile sur photographies : en situation d'évaluation monadique et en situation de comparaison par paire. Ce premier type d'exposition a été choisi parce qu'il permet d'afficher les informations visuelles à une échelle proche de la réalité. Le second type d'exposition a été sélectionné pour sa propension à activer un processus d'évaluation basé sur la comparaison d'éléments visuels, pour son potentiel à produire des mouvements du regard vers des zones locales (comparaison d'items), et pour le doublement du nombre d'objets par rapport à l'exposition monadique.

✓ *Les notes*

Le premier niveau d'observation est l'analyse des évaluations par notation. Les produits sont catégorisés de la même manière dans les deux paradigmes. Pour le produit 207, les deux modalités visuelles sont plus discriminées dans l'évaluation par paire que dans l'évaluation monadique. Cependant pour le produit C6, on observe une dissociation entre la notation et la décision (choix significatif en faveur de la modalité blanche, mais notes moyennes proches). Les notes et les choix ne sont pas directement corrélés car même si les notes sont très similaires entre deux produits, le choix est fortement marqué en faveur de l'un d'eux seulement. Ces résultats sont intéressants dans la mesure où ils montrent que selon le type de produit et d'évaluation, les processus de discrimination peuvent être très différents, et que même si les décisions peuvent refléter une tendance, les notes ne confirment pas celle-ci.

✓ *Conditions expérimentales et comportement oculomoteur*

Au niveau des mouvements du regard, l'objectif de ces expériences est de mettre en évidence des caractéristiques oculomotrices communes à tous les participants effectuant l'évaluation d'un même produit selon une même consigne, pour dégager les spécificités de chacune des conditions expérimentales en termes d'orientation du regard. Cependant, les consignes ne semblent pas avoir directement d'effet sur la sélection visuelle de zones d'informations. Par contre, en terme de dynamique oculomotrice (non pas la prise d'information locale mais la manière dont le regard investit l'image), il apparaît un effet combiné du type de consigne et d'image (organisation spatiale de la scène, et présence/absence d'informations locales) sur le comportement du regard, en terme d'exploration locale (ou « focale ») et d'exploration globale (ou « ambiante »). Ainsi, le produit 207 semble être associé à un traitement visuel plus local que celui de C6, et la consigne de sportivité engendrerait un traitement visuel plus local que la consigne de niveau de gamme. Peu de différences se dégagent entre la consigne « raffinement » et la consigne « niveau de gamme » ; cela pourrait être dû à un faible effet de la consigne sur la prise d'information mais aussi à une forte proximité sémantique entre ces deux consignes. A ce stade de l'étude, le comportement oculomoteur semble être guidé par la manière dont les informations visuelles sont distribuées dans la scène, par le niveau d'abstraction de la consigne, mais aussi par des biais d'observations. En effet, dans les deux configurations d'exposition (par paire et monadique), un fort biais d'observation central est présent. Pour le cas monadique ce biais est positionné au centre de la scène visuelle, et dans les conditions expérimentales de comparaison par paire, ce biais apparaît au centre de chaque modalité visuelle.

✓ *Propriétés temporelles des processus oculomoteurs*

En termes de dynamique temporelle du regard, il semble lors de la comparaison par paire qu'il existe au moins quatre phases : identification, pré-choix, comparaison, stabilisation-décision. On note de plus que, pour la consigne de sportivité, l'habitacle choisi est sur-observé dans la première demi-seconde, et pour toutes les conditions expérimentales l'habitacle choisi est toujours celui qui est (au cumul) le plus longuement observé. Dans l'exposition monadique, il existerait au moins trois phases : prise d'information centrale, exploration excentrique, stabilisation. Au cours de l'exposition, différents processus de guidage du regard seraient activés. Toujours au niveau temporel, on ne dégage pas d'effets des facteurs « consigne », « image » sur les premières zones fixées. Il semble cependant que pour 207, la première figure observée (207-sport) soit celle qui soit choisie.

✓ *Les zones d'intérêt*

Il ne semble pas, dans les paradigmes que nous avons employés, que l'attention visuelle soit dirigée principalement par une activité de collecte d'informations visuelles pour le jugement. L'étude par comparaison par paire met en évidence trois types de zones sur des critères oculomoteurs: les zones fortement observées, tôt dans la séquence et régulièrement, les zones observées secondairement mais avec attention, et les zones délaissées ou juste survolées. Des zones « jumelles » entre les deux modalités visuelles (c'est-à-dire des zones symétriques, par exemple les deux volants ou les deux leviers de vitesse) sont régulièrement classées dans les mêmes groupes, donc bénéficiant du même traitement visuel. Cependant, il n'apparaît pas là non plus de liens évidents entre la consigne et la composition des groupes de zones visualisées. Au niveau des verbalisations, le temps passé sur une zone et l'importance qu'on dit lui accorder dans le jugement sont dissociés : certaines zones sont peu citées et longuement fixées tandis que d'autres sont souvent citées et peu voire jamais fixées. Enfin aucune méthode de comparaison des essais tant sur des critères spatiaux que spatio-temporels ne met en évidence de groupes typiques de comportements oculomoteurs.

✓ *Evaluation d'objets en vision contrainte*

L'expérimentation par vision contrainte complète les observations de la première phase expérimentale. Ainsi, en situation d'exposition monadique, les participants sont capables de décrire l'ensemble du produit depuis le point de vue central. La quantité d'informations visuelles acquises n'est donc pas proportionnelle au nombre de zones observées, et ces informations ne sont pas directement corrélées aux positions du regard. Pour l'exposition par paire, alors qu'il n'apparaît pas être possible de visualiser en détail les deux modalités visuelles depuis le point de vue central (justifiant l'apparition de modes centraux à chaque modalité), un processus de pré-choix de l'habitacle le plus sportif semble se faire sur des critères de luminance. On note l'effet de la « distinctivité » perceptives ou cognitives de certains attributs du produit, attirant l'attention (zone de surbrillance, forme inattendue, configuration anormale de l'objet, etc.). Il s'agit d'un processus attentionnel (de détection plus que d'identification) qui peut induire l'exploration de zones locales (pédalier chromé percé, levier de vitesse brillant, compteurs blancs) et qui se confond avec l'exploration visuelle sur la base de processus sémantiques.

✓ *Evaluation sans contrainte temporelle*

Lors de l'expérimentation avec délai libre, on compare l'évaluation de la sportivité et l'évaluation de l'attractivité sous les formes négatives et positives (la plus vs. la moins). On observe ainsi que la modalité visuelle choisie est sur-fixée dans la première seconde, quelle que soit la valence de la consigne (« la plus sportive » et « la moins sportive »). Cela ne se vérifie pas pour les consignes d'attractivité (modalité préférée vs. la moins appréciée). Un pré-choix orienterait donc le regard dans la première seconde pour la consigne « sportivité » confirmant les observations des expérimentations précédentes. Cette expérimentation met aussi en évidence qu'il existe un biais idiosyncrasique dans le comportement oculomoteur. Cela signifie que la manière d'explorer l'image et de résoudre la consigne

se répète pour les différents essais d'un même participant. En effet, la durée des essais et l'étendue de l'exploration spatiale sont plutôt stables pour un même participant quelle que soit l'image, alors qu'elles sont très variables entre participants. Deux conséquences sont envisageables. La première est que l'influence de la consigne sur le guidage du regard par la consigne peut être très faible par rapport aux processus idiosyncrasiques. La seconde est au niveau de l'enchaînement des processus temporels (identification, évaluation, décision). Puisque que les durées des essais sont très différentes selon les individus, les différentes phases ne sont donc pas synchronisées. Dès lors, ces différences temporelles entre individus peuvent anéantir les effets spécifiques de la consigne lorsqu'ils sont combinés. Ces différences temporelles entre participants peuvent expliquer la sur-observation sur la durée totale d'exposition de la modalité visuelle choisie dans l'expérimentation par paire et à durée contrainte. Alors que Simion observe [simion05] qu'après la décision le regard se répartit également sur les deux modalités visuelles, on note que les participants n'étaient pas invités à justifier leur choix. Or dans notre première expérimentation par paire, cette justification est demandée. On peut ainsi proposer que lorsqu'on demande au participant de justifier un choix, et uniquement si il y a cette demande de justification, celui-ci va, après sa décision, passer du temps à mémoriser la modalité visuelle choisie.

Enfin, cette expérimentation à durée d'exposition libre met en évidence l'existence d'une phase propre à la décision ainsi que des effets de la nature des consignes sur l'orientation du regard. Le regard est donc lié à l'activité cognitive mais ici à un niveau d'orientation macroscopique et non local. L'existence d'un « gaze cascading effect » est confirmée : l'orientation du regard n'est pas uniquement lié à la prise d'information mais aussi à la manifestation d'une attitude (première impression affective), et cette orientation peut biaiser l'évaluation. Ainsi, le lien entre la position du regard et la cognition visuelle n'est pas exclusivement dû aux processus de prise d'information, et l'évaluation n'est pas uniquement dépendante d'une estimation analytique des modalités visuelles.

✓ *Dégager les zones décisionnelles*

Le paradigme des « Bubbles » permet de dégager plusieurs explications sur les processus attentionnels des expérimentations monadiques et par paire. Tout d'abord, les zones de décision pour les deux consignes (caractère sportif et haut de gamme) de 207 sont les mêmes, il y a donc une similarité sémantique des consignes, expliquant la forte similarité entre les quatre conditions expérimentales sur 207 (2 images * 2 consignes). Pour les produits C6, il n'apparaît pas de zones locales de décision (consignes « caractère raffiné » et « haut de gamme »). Cela explique l'exploration visuelle moins locale de ces produits et la faible « spécificité » de chaque modèle de densité de fixations. Cette expérimentation met aussi en évidence un processus de développement de l'expertise au stimulus et à la tâche, influençant la capacité à prendre une décision : plus le participant est expert, plus il est capable de prendre une décision à la fois rapidement et sur des critères ténus. Cela influence directement le comportement oculomoteur car la détection des critères de décision pourra se faire sans fixations locales et la tâche sera temporellement optimisée, donc nécessitant moins de fixations.

✓ *Evaluer les processus attentionnels*

L'évaluation de la contribution de différents facteurs de guidage du regard sur la distribution des points de fixations par condition expérimentale apporte une vision synthétique des processus. La première étude porte sur l'exploration libre (sans consigne) des stimuli employés dans les expérimentations par paire et monadique (dont les résultats convergent). De manière surprenante, la carte sémantique des critères de décision (carte de classification du paradigme des 'Bubbles') contribue significativement à l'explication de la distribution des fixations mesurées sans consigne. Ainsi les interprétations précédentes qui concluent à un lien entre les fixations et les zones locales de décision peuvent être faussées par le fait que même sans recherche d'informations pour les consignes, ces zones sont particulièrement fixées pour des raisons indépendantes de la tâche. Une seconde

observation est que le mode uniforme ne contribue pas aux modèles de distribution des fixations : la structuration spatiale de l'exploration est donc indépendante de la consigne. Enfin, on observe que le modèle de saillance visuelle et de maximisation spatiale de l'information visuelle expliquent plutôt bien certaines distributions. Ainsi, lorsqu'il y a dans l'image des attracteurs attentionnels locaux, comme dans le cas de 207, ce sont ces zones qui vont guider l'exploration visuelle de l'image, même s'il n'y a pas de consignes. Dans le cas où il n'y a pas de zones locales attractives (C6), l'exploration va être plutôt influencée par les facteurs de saillance et de maximisation.

Une seconde étude de l'évaluation des hypothèses de guidage du regard est effectuée pour les évaluations par paire et monadique avec consignes. L'exploration libre étant bien expliquée par la carte sémantique ou par les cartes de saillance/maximisation, on pose l'hypothèse qu'il n'y a pas de grandes différences de distribution du regard entre une situation avec et une situation sans consigne. Nous choisissons alors d'utiliser la carte de densité de l'exploration libre en tant que prédicteur du comportement du regard sur le même stimulus observé avec consigne. Pour l'exposition monadique, le facteur « exploration libre » apparaît comme meilleur prédicteur de l'orientation du regard, puis le facteur centralité (au total 80% des fixations sont ainsi expliquées) alors que le modèle sémantique n'explique qu'entre 0 et 9% des fixations. Le comportement du regard est donc peu différent avec et sans consigne. Une deuxième estimation est effectuée sans le facteur « exploration libre » ni le facteur « biais de centralité ». Les fixations sur le produit 207 sont alors expliquées par la carte sémantique et les fixations sur le produit C6 par les cartes de maximisation/saillance, de la même manière que le cas sans consigne. Pour la situation d'exposition par paire, l'estimation de la contribution des différents facteurs montre que là aussi l'exploration libre prédit bien le comportement avec consigne. Cela signifie que même sans consigne comparative, un processus d'aller et retour entre les deux images est activé. La dynamique comportementale du regard avec consigne n'est donc pas nécessairement liée à la recherche d'informations spécifiques. Tel que pour les cas d'exposition monadique, les distributions des fixations pour 207 sont bien expliquées avec les cartes sémantiques et les distributions des fixations pour C6 le sont avec les cartes de saillance/maximisation. Cependant, malgré ces résultats intéressants, puisque la contribution de chaque facteur dépend des conditions expérimentales, il n'est pas possible de construire un modèle global de prédiction sur la base de ces données.

Discussion/Perspectives

✓ *Bilan*

L'ensemble de ces expérimentations a été conçu dans le but de dégager les relations entre le comportement oculomoteur et la cognition visuelle en situation d'évaluation du produit automobile. L'expérimentation d'évaluation monadique permet d'observer que le comportement oculomoteur est peu impacté par les consignes et que le biais de centralité est fort. On ne note pas de liens directs entre les zones observées et les zones verbalisées. Cependant l'observation multi variée des zones d'intérêt construites *a posteriori* dégage des zones de l'image spécifiquement traitées par le regard lors de la tâche. L'intérêt de cette méthode d'exposition est de mettre en évidence les processus de prise d'information visuelle lorsque les photographies sont proches de l'échelle réelle. L'inconvénient de ce type d'exposition provient de la difficulté à dégager les informations visuelles intégrées au processus d'évaluation. On s'intéresse ensuite à l'exposition par paire, où il est mis en évidence l'existence de liens entre la prise d'information par produit et le choix, ainsi que différents processus influençant la dynamique du regard (dépendants de la présence d'informations locales ou non, de la nature de la consigne générale ou globale, etc.). L'analyse multi variée dégage les zones comparées et les zones délaissées par le regard. Cependant, un biais d'observation centrale à chaque élément de la paire est ici présent, et nous ne mettons pas en évidence de lien entre la consigne et la distribution spatiale des fixations oculaires.

L'exposition par vision contrainte dégage que les informations visuelles perçues ne se limitent pas aux zones fixées, et qu'un processus d'attractivité est activé lorsqu'il y a détection par la vision périphérique d'une différence entre l'objet représenté et le proto-objet mnésique. Le paradigme des « Bubbles » permet d'identifier les zones de décision, qui apparaissent comme pouvant être locales ou globales (207 vs. C6). De plus, deux consignes peuvent se baser sur des zones de décision similaires (sportivité et niveau de gamme), expliquant une partie de la similarité des distributions de fixations entre conditions expérimentales. Le paradigme d'évaluation à durée libre et d'estimation de la préférence montre que la consigne a un effet sur l'orientation du regard, mais plutôt via un rôle de pointeur macroscopique que de capteur local. Le délai de prise de décision apparaît très variable selon les participants (limitant ainsi la pertinence de combiner les différents essais) mais très similaire pour un participant (le comportement oculomoteur ayant donc une forte composante idiosyncrasique) ; cette composante perturbe l'observation des effets des conditions expérimentales sur l'orientation du regard.

Ces résultats permettent de dégager que la définition de notre problématique se décompose en deux parties : celle sur l'étude des liens entre le comportement oculomoteur et l'attention, et celle sur l'étude des liens entre l'attention et les processus d'évaluation. Seule la première est réellement investiguée par nos protocoles expérimentaux, où l'on dégage que le comportement du regard est ici indirectement lié à la consigne. Plusieurs explications peuvent être avancées pour cela. Tout d'abord

une tâche d'évaluation est très différente d'une tâche de recherche de cible, de comptage, de détection ou de lecture (situations classiques de l'emploi de l'oculométrie) car ici les participants ne connaissent pas nécessairement par avance les attributs visuels à analyser pour résoudre la consigne, et car ces attributs visuels peuvent différer d'un individu à l'autre. La nature même de la tâche entraîne ainsi une difficulté à lier les buts et l'orientation du regard. Ensuite, les informations nécessaires à la tâche peuvent être tout autant localement discriminées que globales ; dans cette seconde hypothèse, les zones de fixations ne reflètent alors pas les informations visuelles exploitées pour l'évaluation. C'est le cas semble-t'il pour les consignes de niveau de gamme et du caractère raffiné. Enfin, plusieurs consignes évaluatives peuvent faire appel à des informations locales identiques (similarités sémantiques des consignes), ne permettant donc pas de mettre en évidence les spécificités d'orientation du regard de chaque consigne. Il existe aussi des raisons liées au stimulus expliquant le faible effet de la tâche sur le regard car on note que peut apparaître un fort biais d'observation central à la scène pour l'exposition monadique et central à chaque modalité visuelle pour l'exposition par paire. Un second biais existe pour l'exposition par paire : l'habitacle de droite présente son poste de conduite plus proche du centre de l'image que celui de l'habitacle de gauche, et donc a plus de chance d'attirer le regard. Enfin, il apparaît une désynchronisation temporelle entre les participants dans la résolution de la consigne. Cela explique aussi la difficulté de faire émerger les liens entre la prise d'information visuelle et la tâche évaluative.

On peut synthétiser à partir de nos observations plusieurs paramètres et processus dont les caractéristiques permettent ou non un lien entre le regard et la tâche :

- Les informations visuelles : si elles sont denses, cela implique que le regard acquerra des informations sur plusieurs plans et zones de la scène depuis un seul lieu de fixation, rendant les mouvements oculaires moins utiles, contrairement au cas où elles sont étendues spatialement. Ainsi l'identification des informations que traitent le regard est plus difficile lorsque les informations visuelles sont denses.
- Le positionnement du regard : il peut être positionné sur les zones visuelles traitées par l'attention [Loftus78], mais aussi vers un point de vue spatial optimal [Renninger05a], de manière aléatoire, ou enfin par proches voisins [Sharmin04]. Ainsi, l'inférence de l'orientation de l'attention visuelle à partir de la position du regard n'est pas systématiquement possible.
- L'attention visuelle : elle peut être soumise à des facteurs de guidage de nature physique (saillance) [Itti05], à des facteurs de guidage de nature sémantique (prégnance) [Hammer91], à des facteurs de guidage liés à l'identification (prototype mnésique) [Locher01], à des facteurs comparatifs pour l'exposition par paire (recherche de différences quelles qu'elles soient), orientée pour le traitement de la complexité perceptive et cognitive d'éléments visuels [Leder05], ou encore dirigée vers des zones de forte distinctivité perceptive ou cognitive (innovation, originalité, cohérence, forme inconnue). Le lien entre la tâche évaluative et l'attention peut donc être fortement interféré.
- Attributs évalués : ils peuvent être définis avant, pendant ou après le recueil de l'information visuelle [Russo98], et être de nature uniquement visuelle ou Visu-mnésique. C'est lorsque les attributs sont pré-connus avant exposition et uniquement de nature visuelle (et non mnésiques) que la mesure oculométrique est la plus pertinente.
- Le jugement : il peut être effectué pendant, en fin, ou après exposition, par l'intégration d'attributs visuels de manière analytique ou holistique [Payne93]. Plus le jugement sera analytique plus la position du regard pourra correspondre aux attributs évalués.

La contribution de notre travail à la compréhension des processus d'évaluation des objets manufacturés se fait par des apports méthodologiques liés à l'emploi de l'oculométrie car peu de travaux abordent ces questions dans la littérature. Plusieurs paradigmes expérimentaux sont élaborés afin d'extraire des informations complémentaires par différents protocoles. Nous construisons une boîte à outils de traitements des données provenant de différentes disciplines, et déroulons ces traitements en cascade et en interaction. Nous suggérons ainsi une démarche d'analyse des données par niveau, du global vers le local. Les comportements oculomoteurs sont regroupés afin de limiter la variabilité individuelle. Un premier niveau d'analyse a pour objectif de mettre en évidence des effets macroscopiques sans hypothèse préalable, puis d'identifier des effets à des niveaux toujours plus locaux. Deux types d'observations sont effectués : caractérisation des effets par condition expérimentale, puis catégorisation des conditions sur la base de ces effets. On effectue ainsi une première observation descriptive des données, sans hypothèse préalable sur le comportement du regard, par l'estimation des densités spatiales des fixations *via* les noyaux de Parzen [Parzen62] et les mélanges de gaussiennes [Demster77]. La première méthode démontre son intérêt car étant non paramétrique et permettant un fenêtrage temporel, la seconde car elle permet de caractériser les modèles de distribution des fixations. On dégage ainsi les grandes tendances des comportements du regard, tant spatialement que spatio-temporellement. Le second niveau d'observation porte sur les zones d'intérêt, dont l'établissement *a priori* et *a posteriori* se complète. On montre de plus la complémentarité des algorithmes de « Mean Shift » et de Mélange de Gaussiennes pour l'établissement de groupes de fixations. Les zones d'intérêt sont analysées en terme oculomoteur tout d'abord par une approche mono variée, puis par une approche multi variée (Analyse en Composantes Principales, Analyse Factorielle Discriminante, Classification Ascendante Hiérarchique, Analyse Multidimensionnelle, etc.) permettant de dégager les corrélations entre variables, et des groupements de zone par l'attention visuelle dont elles bénéficient. Enfin, on cherche, par ces méthodes mais sur des variables descriptives non plus de chaque zone mais de chaque séquence oculomotrice, à dégager des groupes d'individus sur la base de leur stratégie de prise d'information visuelle.

Une démarche d'aide à la conception des expérimentations sur l'évaluation visuelle de produits manufacturés est ainsi proposée.

- *Evaluer l'apport de l'oculométrie*

Observer comment s'oriente l'attention visuelle lorsque des produits sont évalués permet d'aider à la mise au point de ces produits en phase de conception. Cela peut aussi permettre de comprendre les différentes attitudes des clients et donc d'orienter la conception en fonction de cibles. Cependant, la mesure oculomotrice est intéressante dans les situations où les informations visuelles étudiées sont locales et discriminées spatialement, lorsqu'il peut exister des stratégies homogènes entre participants (experts vs. novices par exemple), et lorsque l'attention est plutôt de nature séquentielle que parallèle. L'intérêt de la mesure du regard est donc dépendant

- du caractère local et distinct des informations visuelles pertinentes
- du caractère homogène inter-participant de la prise d'information
- du caractère heuristique ou analytique de la consigne

Afin d'exploiter les mesures oculomotrices, il est donc nécessaire de bien définir le type d'informations que l'on souhaite relever (consigne) et les modalités visuelles à évaluer (stimuli) pour considérer si le comportement oculomoteur peut permettre d'inférer les informations recherchées. Ainsi, une première étape est de définir si les différentes informations visuelles à discriminer peuvent être appréhendées ou non depuis une seule et même position du regard. Une seconde étape est de vérifier que l'évaluation se base bien sur des critères visuels, et que ces informations visuelles sont appréhendables localement. Une troisième étape est de tester si les informations visuelles traitées

sont effectivement représentatives d'une stratégie ou d'un processus attentionnel stable, et non de processus idiosyncrasiques ou non orientés.

- *Orienter les processus de l'attention visuelle vers la résolution de la tâche*
- Améliorer la familiarité

Afin de dégager des effets robustes, il peut être pertinent de cadrer la familiarité du participant vis à vis de l'objet et de favoriser une prise d'information locale. Pour cela, il semble intéressant d'exploiter les effets de la familiarisation au stimulus, qui, plutôt qu'un biais, pourrait permettre d'optimiser la prise d'information propre à une consigne. C'est pourquoi dans le paradigme du « Repeated Evaluation Technic » [Carbon05] l'exposition se fait en deux phases, une première qui sera associée à une investigation de l'image pour l'apprentissage (catégorisation) et l'attitude première (appréciation), et une seconde phase pour l'évaluation proprement dite, dans laquelle les processus attentionnels seront potentiellement plus orientés vers la recherche d'informations propres à la consigne.

- Utiliser l'interactivité

L'utilisation de l'interactivité dans les paradigmes exploitant la mesure du comportement oculaire peut être d'une réelle plus-value. En effet, l'une des difficultés de nos situations expérimentales est que le regard est soumis à des processus parallèles et non séquentiels de traitement de l'information. Si le participant interagit avec le stimulus, alors son attention va être séquentialisée par l'action [Duchowski03]. Les informations visuelles traitées et les processus cognitifs engagés vont se manifester de manière plus formelle et moins ambiguë. On peut ainsi proposer un paradigme où le participant manipule lui-même une caméra dans une scène virtuelle, avec une fonction de déplacement et une fonction de grossissement de l'image. En mesurant les mouvements de caméra et les positions du regard, on pourrait dégager les phases de recherche active d'informations des phases d'intégration cognitive, les phases d'évaluation globale des phases d'exploration locale, etc.

- Favoriser la dynamique du regard

Le biais de centralité et donc de prise d'information globale peut être influencé par le dispositif d'affichage. Ainsi, plutôt que de présenter les habitacles sur un même écran, on peut proposer qu'un système multi écrans gère l'affichage dans le cas de l'évaluation par paire, permettant d'afficher les produits à une plus grande échelle, donc avec une meilleure discrimination des zones locales, et engendrant des orientations plus manifestes du regard vers la modalité visuelle observée. On peut de plus proposer que les différentes modalités visuelles soient disposées verticalement et non horizontalement dans la scène, pour éviter qu'un poste de conduite soit plus proche du centre de la scène que les autres. Une autre évolution qui peut être appliquée tant sur le cas par paire que monadique consisterait à décentrer aléatoirement les positions à la fois du point de fixation original, et des modalités visuelles dans la scène. Une troisième proposition serait de faire apparaître l'image progressivement. L'idée est d'engendrer une dynamique de l'orientation du regard par sa synchronisation avec la dynamique d'apparition des informations visuelles. Ainsi, si l'on fait apparaître l'image par plage de fréquences, en affichant progressivement des plages de fréquences basses puis hautes, les informations de contour (basses fréquences) vont d'abord apparaître, orientant le regard vers ces zones, puis lorsque des détails (hautes fréquences) vont survenir en périphérie, alors le regard s'y dirigera, perpétuant le décentrage initié.

- *Mesurer le regard*
- Choisir les durées d'exposition

Au niveau des durées d'exposition, il semble qu'il soit plus pertinent sur ce type de tâches de laisser le temps de réponse libre, le participant mettant lui-même fin à l'essai, du fait de la

désynchronisation temporelle des processus de prise de décision et de la sur-observation de la modalité visuelle choisie.

- Rendre robuste les mesures

Afin d'améliorer le traitement des mesures oculométriques et leur significativité, on suggère d'obtenir au moins 12 essais de qualité suffisante. Il serait donc pertinent de définir un indice de qualité des mesures, permettant d'accepter les essais ou de les rejeter. Cet indice pourrait se baser sur une mesure de la précision sur un point référence en calculant la dérive avant et après la tâche, ainsi que sur le temps de tracking perdu ([Hornof02] [Salvucci01]). Enfin, on a pu constater que la précision des mesures était meilleure avec un oculomètre à casque par rapport à un oculomètre déporté, et avec une mentonnière plutôt que sans. Il sera nécessaire d'évaluer au cas par cas si la dégradation du caractère écologique de la mesure sera fortement perturbatrice des processus attentionnels ou non.

- *Traiter les données*

Plusieurs évolutions des traitements de données oculométriques peuvent aussi être envisagées. De nouvelles méthodes de classification des zones d'intérêt en fonction du comportement oculomoteur peuvent être étudiées, ainsi que pour la classification des essais, en intégrant notamment la présence d'effets non linéaires. Les algorithmes de classifications auto-organisées (« Self Organized Map » [Kohonen97]) et d'Analyse en Composantes Indépendantes [Hyvärinen01] semblent intéressants, car les analyses multi variées que l'on a employées ne quantifient que des effets linéaires sur des dimensions pouvant être liées. Des travaux de Salojarvi [Salojarvi03] en démontrent l'intérêt. Les réseaux de neurones, systèmes Bayésiens ([Baddeley05], [Itti06b]) et Chaînes de Markov [Salvucci01] pourraient permettre d'aller plus loin dans l'identification des fixations et saccades, dans la reconnaissance de motifs comportementaux répétitifs, dans l'évaluation de l'homogénéité intra-conditions expérimentales, et dans la détermination des facteurs prédisant les zones de fixation. Au niveau de la modélisation des facteurs d'orientation du regard, le développement d'algorithmes intégratifs sur la base des cartes de saillances et de cartes sémantiques semble prometteur et donc une direction de recherche à privilégier. Enfin, le concept de carte de maximisation constitue une direction de recherche à développer et à enrichir lors de l'emploi de scènes réelles complexes telles que celles d'objets manufacturés, car il peut permettre une bonne prédiction des comportements du regard complétant les cartes de saillance et cartes sémantiques.

- ✓ *Perspectives*

Nous présentons dans cette fin de section les perspectives tout d'abord selon les axes théoriques que l'on entrevoit dans la continuité de nos travaux, puis ensuite par les champs applicatifs qui pourraient exploiter les méthodes développées.

Afin d'enrichir les réflexions théoriques abordées dans ce travail, les paradigmes de « Rapid Serial Visual Display » (exposition rapide séquentielle [Williams98]) pourraient permettre d'identifier en fonction des durées d'exposition les orientations du regard (observation de l'attention), le type d'informations visuelles encodées (observations du traitement des informations), et les résultats évaluatifs (observation de l'intégration de ces informations). Cela apportera de nouveaux indices quant à la manière dont les informations visuelles sont acquises et influencent l'évaluation. Un deuxième axe à développer est l'étude des biais d'évaluation liés à l'aisance au traitement des informations (perceptivement et cognitivement) pour mieux comprendre les processus de formation d'impression.

Afin d'aller plus loin dans la compréhension des processus d'évaluation visuelle, il serait nécessaire d'étudier les relations entre l'attention visuelle et l'évaluation par les champs de la cognition visuelle et de la psychologie sociale et de décision. L'idée serait d'approfondir la compréhension des processus

cognitifs sur lesquels le jugement visuel s'appuie, ainsi que les heuristiques et biais de décision, les liens entre le jugement et les informations perceptives et mnésiques, et les effets de la pré-connaissance ou de l'expertise sur l'évaluation. Cela permettrait de lier nos travaux à ceux portant sur la catégorisation cognitive des objets et sur l'inférence des traits et attributs des produits, afin d'aboutir à une compréhension globale de la chaîne perceptive et cognitive activée lors de l'évaluation des produits manufacturés.

L'étude des mouvements du regard en situation d'évaluation de produits réels est aussi à envisager. En effet, les informations de volume et de profondeur sont peu présentes sur images et la dynamique de luminance ainsi que la plage de chrominance sont restreintes sur photographies par rapport au réel. De plus le champ visuel est plus grand en réel que sur image, et enfin l'implication du participant dans la tâche est plus forte lorsqu'il se trouve en situation écologique [Ijsselsteijn01]. On peut ainsi s'interroger sur la convergence entre l'évaluation d'un produit photographié et son évaluation réelle. Du fait des travaux contemporains sur la discrimination binoculaire et sur la stéréovision, associé à l'amélioration des techniques de mesure du regard permettant de relever la distance entre l'objet observé et l'observateur, l'étude du comportement oculomoteur en situation réelle peut permettre de mieux dégager les informations visuelles vers lesquelles s'oriente l'attention, et ensuite intégrées pour le jugement. En comparant les comportements oculomoteurs en situation d'évaluation d'un stimulus réel et de son image, les lacunes informationnelles de la situation 2D pourraient être mises en évidence.

Au niveau du déploiement applicatif de nos travaux, il pourrait être intéressant d'étudier les processus oculomoteurs et d'évaluation lorsque le participant est face à plusieurs écrans présentant soit différents axes de vue d'un même produit, soit plusieurs produits. On peut alors tenter de dégager par les relevés oculomoteurs les stratégies de sélection des informations visuelles, les points de vue favorisés, les liens entre la construction de la décision et les modalités visuelles observées, etc. Puisque dans ces situations les différentes modalités visuelles sont bien discriminées spatialement, le regard devrait manifester l'orientation de l'attention visuelle. Une autre exploitation de l'oculométrie se base sur le traitement en temps réel de la position du regard. Par exemple, en associant la mesure du comportement du regard à d'autres relevés tels que le choix d'axe de vues, un système de détection automatique des préférences peut être élaboré. Par exemple, un individu examine en détail les textures des sièges ou compare les combinés de deux habitacles ; la détection de cette intention va alors servir à configurer une deuxième phase évaluative, dans laquelle plusieurs versions des éléments vers lesquels s'est portée l'attention sont proposées. La détection des intérêts sans nécessité de faire un effort mnésique ou de verbalisation pourrait engendrer une forte aisance de tâche, favorable à la mise en évidence des critères de décision et d'élaboration du jugement. L'orientation du regard en tant que pointeur manifestant l'intérêt et la préférence à un niveau macroscopique peut aussi être exploitée. Une autre application serait d'exploiter le regard en tant que système d'interface afin de paramétrer en temps réel la scène exposée en fonction de la position du regard. Le regard pourrait par exemple servir à naviguer dans la scène ou à afficher des informations contextuelles complémentaires sur les attributs du produit. Cela pourrait améliorer l'implication du participant dans sa tâche par un processus de synchronisation entre le système technique et le système cognitif, par l'optimisation de la vitesse et de l'aisance de navigation, ainsi que par l'amélioration de la crédibilité médiatique [Holbrook82]. La multi-modalité nous semble aussi être une perspective majeure d'exploitation de la mesure de la position du regard. L'étude des interactions multimodales (acoustiques & visuelles par exemple) peut s'appuyer sur le comportement oculaire pour étudier le lien entre les flux visuels, les flux acoustiques et le résultat de l'intégration de ces deux flux. De même avec le sens haptique, où il est intéressant d'observer si les informations visuelles combinées par exemple à des informations haptiques texturales modifient la perception globale ou l'évaluation [Lécuyer01]. Dans des situations d'immersion virtuelle, la mesure du regard vient aider à la compréhension des liens entre perception de la dynamique des mouvements et événements visuels.

Alors que la compréhension des processus d'évaluation des produits par les consommateurs est toujours plus stratégique dans l'industrie et que les modèles d'attention visuelle et de prise de décision sont de plus en plus sophistiqués, notre travail éclaire une petite parcelle du champ d'étude pluridisciplinaire des mécanismes d'évaluation visuelle des objets manufacturés. En se positionnant dans le domaine général des sciences cognitives, notre travail ne permet pas de répondre à toutes les problématiques évoquées ni de proposer de nouveaux modèles attentionnels, mais par une vision unifiée des méthodes d'analyse et de traitement des données, va aider à caractériser et catégoriser les comportements oculomoteurs, pour une mise en évidence des processus de cognition visuelle.

Mots-clés : traitement de données, oculométrie, mouvements des yeux, perception visuelle, cognition visuelle, évaluation, jugement, produit manufacturé, automobile

L'intégration des facteurs humains dans la conception et le développement industriel est aujourd'hui un enjeu majeur de l'innovation technologique, et plus particulièrement pour l'automobile. L'étude des processus d'évaluation des produits manufacturés permet ainsi d'apporter des informations clés sur la manière dont le produit va être perçu puis jugé. Par son approche pluridisciplinaire à la confluence des sciences de l'ingénieur et des sciences cognitives, notre travail contribue à la compréhension des processus de la cognition visuelle et de formation d'impression par l'usage de l'oculométrie. Ainsi, les comportements oculomoteurs mesurés en situation d'évaluation visuelle du produit automobile vont être caractérisés et catégorisés par des méthodes statistiques empruntées au traitement des images, à la biologie, à l'agronomie ou la psychophysique. Les liens entre l'orientation du regard, l'attention visuelle et l'intégration cognitive des informations visuelles pour l'évaluation seront ensuite étudiés. Plusieurs paradigmes expérimentaux sont conçus : évaluation monadique ou par paire, à durée d'exposition libre ou limitée, depuis des positions d'observations libres ou contraintes. Il s'avère que les processus qui guident l'orientation du regard lors de l'évaluation visuelle de produits photographiés, tels que les habitacles automobiles, sont nombreux et imbriqués de manière complexe tant spatialement que temporellement. Ce travail développe non seulement un ensemble de protocoles et de préconisations d'usage de l'oculométrie, mais aussi une démarche d'analyse de ces données par des traitements en cascade, afin de mettre en évidence le traitement visuel des zones d'intérêt et les effets des consignes. Finalement, un ensemble de méthodes est proposé pour estimer les contributions relatives des différents processus attentionnels sur la structuration des distributions de fixations.

Key-words: data analysis, eye movement, visual perception, visual cognition, assessment, evaluation, judgment, manufactured objects, automotive

Today, Human factors are highly strategic in the industrial process of design and technical innovation, and have to be considered in particular for automotive industry. Especially, the consumer's behavior studies are very useful for a better understanding of a judgment process. In that way, our work deals with engineering and cognitive sciences. It contributes by the use of eye movements to analyze and model visual cognition and assessment process. At first, the eye movements are characterized and classified with statistical tools coming from image processing, biology, agronomy or psychophysics. Then, we observe links between gaze orientation, visual attention, and cognitive integration of visual information. Different experimental paradigms are built: monadic and pair-wise evaluation, free or fixed exposition duration, constrained or free point of view. In the context of automotives evaluations, gaze behavior seems to be influenced by a lot of factors, organized in a complex manner. That's why we suggest new protocols and recommendations for the use of eye-movement measures, in these complex situations. A "statistical eye movement analysis toolbox" and a "cascading treatment methodology" are built for observation of visual behavior on areas of interest and for visual information integration. Finally, we suggest new methods to evaluate the different attention process effects on eye movements.

Références

- [Abrams92] Abrams, R.A. (1992). Planning and producing saccadic eye movements. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (p.66); New York: Springer-Verlag
- [Ahn00] Ahn, W. & Dennis, M. J. (2000). Dissociation between categorization and similarity judgment: differential effect of causal status on feature weights. In Hahn, U and Ramscar, M. (Eds.) *Similarity and categorization*. Oxford: Oxford University Press.
- [Aitkin80] Aitkin, M., Tunnicliffe Wilson, G. (1980). Mixture Models, Outliers and the EM Algorithm. *Technometrics*, 22, 325-332.
- [Ajzen85] Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action-control: From cognition to behavior* (pp. 11-39). Heidelberg: Springer.
- [Alba00] Alba, J.W., Hutchinson, J.W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of consumer research*, 27 (September), 123-156.
- [Alba87] Alba, J., Hutchinson W. (1987) "Dimensions of Consumer Expertise", *Journal of Consumer Research*
- [Altmann04] Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the "blank screen paradigm". *Cognition*, 93, B79-B87.).
- [Antes81] Antes, J. R., & Penland, J. G. (1981). Picture context effects on eye movement patterns. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 157–170). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- [Ariely01] Ariely, Dan. (2001) Seeing Sets: Representation by Statistical Properties. *Psychological Science* 12 (2), 157–162.
- [Arnheim54] Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye*. Berkeley: University of California Press.
- [Baccino01] Baccino, T., Colombi, T. (2001). Les interactions Homme-système : perspectives et recherches psycho ergonomiques, pp127-148, Paris :Hermès ; (in A. VomHofe (Ed.)
- [Baccino02] Baccino, T. (2002). Oculométrie Cognitive, In G. Tiberghien (Ed.), *Dictionnaire des Sciences Cognitives*, pp. 100-101, Paris :Armand Colin.
- [Baddeley05] Baddeley, R. J., Tatler, B. W. (2005). High frequency edges (but not contrast) predict where we fixate - A Bayesian system identification analysis. *Vision Research* 46 (2006) 2824-2833
- [Bahill75] Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology*, 14, 468–469.
- [Bar03] Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience*, 15, 600-609.
- [Bar04] Bar, M. (2004). Visual objects in context. *Nature REviews Neuroscience*, 5, 617-629.
- [Barsalou85] Barsalou, L. W. (1985). Ideas, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of experimental psychology : Learning, Memory and cognition*, 11(4), 629-654.
- [Barth06] Barth, E., Dorr, M., Böhme, M., Gegenfurtner, K. R., Martinez, T. (2006). Guiding eye movements for better communication and augmented vision. In *Perception and Interactive Technologies*, volume 4021 of *Lecture Notes in Artificial Intelligence*, pages 1-8. Springer 2006.
- [Barton89] Barton, M.E., Komatsu, L.K. (1989). Defining features of natural kinds and artifacts, *Journal of psycholinguistic research*, 1989, vol. 18, no5, pp. 433-447
- [Beatty87] Beatty S.E. et Smith S.M. (1987), External Search Effort: An Investigation Across Several product Categories, *Journal of Consumer Research*, 14, 1, 83-95
- [Berlyne70] Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception and Psychophysics*, 8(5-A), 279–286.
- [Berlyne71] Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. Appleton-Century-Crofts, New York
- [Bertamini05] Bertamini, M. Jones, L. A., Sponner, A., Hecht, H. (2005). Boundary extensions : the role of magnification, object size, context and binocular information. *Journal of Experimental Psychology Human : Perception and Performance*, 2005, Vol. 31, No. 6, 1288–1307
- [Biederman06] Biederman, Irving & Vessel. (2006) *Perceptual Pleasure and the Brain*. *American Scientist*. 94(3), 247-253.
- [Biederman72] Biederman, I. (1972). Perceiving real world scene. *Science*, 177, 77-80
- [Biederman81] Biederman, I. (1981). On the semantics of a glance at a scene. In : M. Kubovy, and J. R. Pomerantz (ed.) *Perceptual organization*. Hillsdale, NJ : Erlbaum, pp.213-263.
- [Biederman82] Biederman, I. (1982). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–254). Hillsdale, NJ: Erlbaum.

28. [Biederman82b] Biederman, I. et al. (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cognit. Psychol.* 14, 143–177
29. [Biederman83] Biederman, I., Teitelbaum, R. C., & Mezzanotte, R. J.(1983). Scene perception: A failure to find a benefit from prior expectancy or familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 411–429.
30. [Biederman98] Biederman, I. (1998). Aspects and extension of a theory of human image understanding. In *Computational Processes in Human Vision: An Interdisciplinary Perspective*, Z. Pylyshyn (Ed.), New Jersey: Ablex Publishing Corporation.
31. [Biernacki02] Biernacki, C. Celeux, G. and Govaert, G. (2002) "Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models". *Computational Statistics and Data Analysis*, (to appear).
32. [Biernacki06] Biernacki, C., Celeux, G., Govaert, G., Langrognet, F. (2006.)Model-Based Cluster and Discriminant Analysis with the MIXMOD Software. *Computational Statistics and Data Analysis*, vol. 51/2, pp. 587-600.
33. [Bisley03] Bisley, J.W. and Goldberg, M.E. (2003) Neuronal activity in the lateral intraparietal area and spatial attention. *Science* 299, 81–86
34. [Bisseret99] Bisseret, A ; Sebillotte, S. & Falzon, P. *Techniques pratiques pour l'étude des activités expertes*, Toulouse, Octarès, 1999
35. [Boloix05] Boloix, E. (2005). Influence de la tâche sur le traitement des scènes visuelles complexes : une modélisation des niveaux de représentation. Thèse de doctorat, Université de Provence.
36. [Bornstein89] Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265–289.
37. [Bowmaker80] Bowmaker J.K. , Dartnall H.J.A. (1980). Visual pigments of rods and cones in a human retina. *J. Physiol.* 298: pp501-511.
38. [Boyce89] Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 556–566.
39. [Bradley52] Bradley, R.A., Terry, M.E. (1952). The rank analysis of incomplete block designs. I. The method of paired comparisons" *Biometrika* , 39 (1952)
40. [Brandt45] H. F. Brandt (1945) *The Psychology of Seeing*, Philosophical Library, New York.
41. [Brandt97] Brandt, S. A., Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, 9, 27–38
42. [Braun99] Braun, K.A. (1999). Psotexperience advertising effects on consumer memory. *Journal of consumer research*, 25 (march), 319-334
43. [Brenner96] Brenner, L.A., Koehler, D.J., Tversky, A. (1996). On the Evaluation of One-Sided Evidence ; *Journal of Behavioral Decision Making*, 9 (March), 59–70.
44. [Bridgeman75] Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15(6), 719–722.
45. [Brockmole05] Brockmole, J. R., & Henderson, J. M (2005). Prioritization of new objects in realworld scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 857-868.
46. [Brockmole05b] Brockmole, J.R., Henderson, J.M. (2005). Object Appearance, disappearance, and attention prioritization in real world scenes. *Psychonomic Bulletin & Review*, 200(, 12(6), 1061-1067
47. [Brockmole06] Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, 13, 99-108.
48. [Brockmole06b]]Brockmole, J. R., Castelhani, M., S., Henderson, J. M. (2006). Contextual cueing in naturalistic scenes : Global and local contexts. *Journal of experimental psychology: learning, memory and cognition* 2006, Vol 32, N°4, 699-706.
49. [Broniarczy94] Broniarczy, S.M., Alba, J.W. (1994). Theory versus data in prediction and correlation tasks. *Organizational behavior and human decision process*, 57 (january), 117-139.
50. [Bruce90] Bruce, V., & Green, P.R. (1990). *Visual Perception: physiology, psychology and ecology* (2nb edition). Lawrence Erlbaum Associates Ltd., Hove, UK
51. [Buswell35] G. T. Buswell, *How People Look at Pictures: A Study of The Psychology of Perception in Art*, The University of Chicago Press, Chicago, 1935
52. [Cajal11] Cajal, S.R.Y.(1911), *Histologie Du Système Nerveux de l'Homme et Des Vertébrés*, Maloine, Paris
53. [Carbon05] Carbon, C.C., Leder, H. (2005). The Repeated Evaluation Technique (RET). A Method to Capture Dynamic Effects of Innovativeness and Attractiveness. *Applied Cognitive Psychology*. 19: 587-601 (2005)
54. [Carpenter83] Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275-307). New York: Academic Press

55. [Carver81] Carver, C.S., Scheier, M.F. (1981). *Attention and self regulation : a control theory approach to human behavior*, NewYork : Springer-Verlag
56. [Caspi04] Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 13086–13090.
57. [Castelhana03] Castelhana, M. S., & Henderson, J. M. (2003). Flashing scenes and moving windows: An effect of initial scene gist on eye movements
58. [Castelhana05] Castelhana, M. S., & Henderson, J. M., (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, 12, 1017-1040.
59. [Cavanagh01] Cavanagh, Patrick (2001) Seeing the forest but not the trees ; *Nature Neuroscience* 4, 673 - 674 (2001)
60. [Cellerino03] Cellerino, A. (2003). "Psychobiology of facial attractiveness." *J Endocrinol Invest* 26(3 Suppl): 45-8.
61. [Chauvin03] Chauvin, A. (2003). Perception des scènes naturelles: étude et simulation du rôle de l'amplitude, de la phase et de la saillance dans la catégorisation et l'exploration des scènes naturelles. *These de Doctorat, Université pierre mendès-france, Grenoble*
62. [Cheng95] Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790-799, Aug. 1995.
63. [Chun98] Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28-71
64. [Churchland94] Churchland, P. S., Ramachandran, V. S. and Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch and J. L. Davis (Eds.) *Large Scale Neuronal Theories of the Brain*, pp. 23-60, Cambridge, MA: MIT Press.
65. [Collins69] Collins, A.M., Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8, 240-248.
66. [Collins70] Collins, A.M., Quillian, M.R. (1970). Facilitating retrieval from semantic memory : the effect of repeated part of inference. *Acta Psychologica*, 33, 304-314.
67. [Collins72] Collins, A.M., Quillian, M.R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.). *Cognition in learning and memory*. New York, Wiley.
68. [Cordier02] Cordier, F. (2002). Les theories conceptuelles. In F. Cordier & J. François (Eds.), *Catégorisation et langage*. Paris Hermès.
69. [Cordier93] Cordier, F. (1993). Les representations cognitive privilégiées. Typicalité et niveaux de base. Lille, Presses Universitaires de Lille.
70. [Cornsweet73] Cornsweet, T. N., Crane, H. D. (1973); Accurate two-dimensionnal eye tracker using first and fourth Purkinje images. *Journal of the Optical Society of America*, 63, 921.
71. [Craik72] Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
72. [Crandall73] Crandall, J. E., Montgomery, V. E., Rees, W.W. (1973). "Mere" exposure versus familiarity, with implications for response competition and expectancy arousal hypotheses. *Journal of General Psychology*, 88(1), 105–120.
73. [Currie00] Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics*, 62(4), 673–683.
74. [Dasgupta99] Dasgupta, S. (1999). "Learning Mixtures of Gaussians". *Proc. of Symposium on Foundations of Computer Science (FOCS)*.
75. [Davenport04] Davenport, J. L., Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological science*, 15, 559-564.
76. [DeGræf90] De Græf, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, 52, 317-329.
77. [DeGraef98] De Graef, P. (1998). Prefixational object perception in scenes : objects popping out of schemas. In : G. Underwood (ed.). *Eye guidance in reading and scene perception*. Oxford : Elsevier, pp.315-338.
78. [Dempster77] Dempster, A. P., Laird, N. M. Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39, 1-38.
79. [Denisova06] Denisova K, Singh M, Kowler E, 2006, "The role of part structure in the perceptual localization of a shape" *Perception* 35(8) 1073 – 1087
80. [Deubel96] Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12), 1812–1837.
81. [Dhar00] Dhar, R., Nowlis, S.M., Sherman, S.J. (2000). Trying hard or hardly trying: An analysis of context effects in choice. *Journal of Consumer Psychology*, 9 (4), 189-200.

82. [Dixon48] Dixon, W. R., Jr. (1948). Studies of the eye-movements in reading of university professors and graduate students. Microfilm Abstracts. ; eurich 1933a - Eurich, A. C. (1933a). Fourth-and fifth-grade standards for photographic eye-movement records. *Journal of Genetic Psychology*
83. [Dodge1900] Dodge, R. (1900). Visual perception during eye movements. *Psychological review*, 7, 454-465.
84. [Dodge1901] Dodge, R., Cline, T. S. (1901). The angle velocity of eye movements. *Psychological review*, 8, 145-157.
85. [Dompnier06] Dompnier, Lucille (2006) Le jugement de qualité d'un objet manufacturé : le réseau conceptuel sous-jacent. Thèse de doctorat, Université de Poitiers, Psa Peugeot Citroen.
86. [Donnet03] Donnet, S. (2003). Contrôle des saccades vers des cibles dans l'espace tridimensionnel : étude en laboratoire et en situation naturelle de conduite automobile. Thèse de doctorat, Université Paris 5
87. [Dorr05] Dorr, M., Martinetz, T., Bohme, M., Barth, E. (2005) Visibility of temporal blur on a gaze contingent display; ACM vol. 95, proceedings APGV
88. [Downing88] Downing, C. J. (1988). Expectancy and visual-spatial attention : Effects on perceptual quality. *Journal of experimental psychology : Human Perception and performance*, 14, 188-202.
89. [Duchowski01] Duchowski, A., Mdelin, E., gramopadhye, A., Melloy, B., & Nair, S. (2001). Binocular Eye tracking in VR for Visual Inspection Training. In *Virtual Reality Software & Technology (VRST)*. Banff, AB, Canada
90. [Duchowski03] Duchowski, A. (2003). *Eye Tracking Methodology, Theory and Practice*. Ed. Springer
91. [Emery00] Emery, N. J. (2000). "The eyes have it: the neuroethology, function and evolution of social gaze." *Neuroscience And Biobehavioral Reviews* 24(6): 581-604.
92. [Engel97] Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, 7, 181-192.
93. [Epstein98] Epstein, R. & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601.
94. [Etcoff99] Etcoff, N. (1999). *Survival of the prettiest: The science of beauty*. New York, NY: Anchor Books/ Doubleday.
95. [Falk86] Falk, D. S., Brill, D. R., Stork, D. G. (1986) *Seeing the Light: Optics in Nature, Photography, Color Vision and Holography*. John Wiley & Sons Edt.
96. [Fantz64] Fantz, R. L. (1964). "Visual Experience In Infants - Decreased Attention To Familiar Patterns Relative To Novel Ones." *Science* 146(364):
97. [Fazendeiro02] Fazendeiro, T. A., Winkelman, P., Luo, C., & Lorah, C. (2002). False recognition across meaning, language and stimulus format: Conceptual relatedness and feelings of familiarity. *Memory and Cognition*.
98. [Fecteau06] Fecteau, J., Munoz, D. (2006). Saliency, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, Volume 10, Issue 8, Pages 382-390
99. [Fei-Fei04] Fei-Fei, L., Fergus, R., & Perona, P. (2004). Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In *IEEE International workshop on Generative-Model Based Vision*, in conjunction with CVPR'05.
100. [Fei-Fei05] Fei-Fei, L., & Perona, P. (2005). A Bayesian hierarchy model for learning natural scene categories. In *IEEE Computer Vision and Pattern Recognition*. San Diego, USA. VII: 524-531.
101. [Fei-Fei07] Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1):10, 1-29, <http://journalofvision.org/7/1/10/>.
102. [Feldman88] Feldman, J.M., Lynch, J. G. (1988). Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (August), 421-435.
103. [Ferreira86] Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368
104. [Findlay01] Findlay, J. M., Gilchrist, I. D. (2001). *Visual Attention : the active vision perspective*
105. [Findlay97] Findlay, J.M.(1997) saccade target selection during visual search. *Vision research*, 37, 617-631
106. [Findlay99] Findlay, J. M. and Walker, R. (1999). A model of saccadic eye movement generation based on parallel processing and competitive inhibition. *Behav. And Brain Sci.*, 22:661-674
107. [Fishbein75] Fishbein, M. and Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley, Reading, MA.
108. [Fitzsimons01] Fitzsimons G.J., Shiv, B. (2001). "Nonconscious and Contaminative Effects of Hypothetical Questions on Subsequent Decision Making ; *Journal of Consumer Research*, 28 (September) 224-238.
109. [Forgy65] Forgy, E. (1965) Cluster analysis of multivariate data : Efficiency vs. interpretability of classifications. *Biometrics*, page 21:768, 1965.
110. [Fraley98] Fraley, C., Raftery, A. E. (1998). How many clusters? Which clustering method? - Answers via Model-Based Cluster Analysis. Technical Report no. 329, Department of Statistics, University of Washington, February 1998. *Computer Journal* 41:578-588
111. [Freund96] Freund, T., Kruglanski, A., & Bar-Tal, D. (1996). Motivational effects in the "mere-exposure" paradigm. *European Journal of Social Psychology*, 26, 479-499.

112. [Friedman79] Friedman, A. (1979). Framing pictures : the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental psychology: General*, 108, 316-355.
113. [Friesen03] Friesen, C. K. and A. Kingstone (2003). "Abrupt onsets and gaze direction cues trigger independent reflexive attentional effects." *Cognition* 87(1): B1-10.
114. [Frost76] Frost, D., & Pöppel, E. (1976). Different programming modes of human saccadic eye-movements as a function of stimulus eccentricity—indications of a functional subdivision of visual-Weld. *Biological Cybernetics*, 23(1), 39–48.
115. [Fukunaga75] Fukunaga, K. Hostetler, L.D. (1975) "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Trans. Information Theory*, vol. 21, pp. 32-40, 1975
116. [Garcia00] Garcia-Marques, T. Mackie, D.M. (2000). The positive feeling of familiarity: mood as an information processing regulation mechanism. In J. Forgas and H. Bless (Eds.), *The message within: The role of subjective experiences in social cognition and behavior* (pp. 240-261). Philadelphia: Psychology Press.
117. [Germeys01] Germeys, F., & d'Ydewalle, G. (2001). Revisiting scene primes for object locations. *The Quarterly Journal of Experimental Psychology*, 54A, 683-693.
118. [Gibson77] Gibson, J. J. (1977). *The Theory of Affordances*. In *Perceiving, Acting, and Knowing*, Eds. Robert Shaw and John Bransford,
119. [Gibson79] Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
120. [Goldberg98] Goldberg, J.H., and Kotval, X.P. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of industrial Ergonomics*. 24, 1998, pp.631-645
121. [Gosselin01] Gosselin, F., Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research* 41 (2001) 2261–2271
122. [Greene06] Greene, M.R., & Oliva, A. (2006). Natural scene categorization from the conjunction of ecological global properties. *Proceeding of the Cognitive Science Meeting, Vancouver, August 2006*.
123. [Grill-Spector05] Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, 16, 152–160
124. [Grimes96] Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception: Vancouver studies in cognitive science* (Vol. 2, pp. 89–110). New York: Oxford University Press.
125. [Groner84] Groner, R., Walder, F., & Groner, M. (1984). Looking at faces: Local and global aspects of scanpaths. in A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 523-533). Amsterdam: Elsevier.
126. [Guerin00] Guerin-Dugue, A., Oliva, A. (2000). Classification of scene photographs from local orientations features. *Pattern Recognition Letters*, 21:1135–1140.
127. [Guyader04] Guyader, N. (2004). Perception de scenes et caractérisation d'objets en fonction du contexte : taille, orientation, perspective. Thèse de doctorat soutenu au Laboratoire des Images et signaux, INPG, Grenoble.
128. [Halberstadt03] Halberstadt, J., Rhodes, G. (2003). "It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive." *Psychonomic Bulletin & Review* 10(1): 149-156.
129. [Hammer89] Hammer, N. and Lengyel, S. (1989). Do people perceive what the design expresses? In Vi-hma, S. (ed.). *Semantic Visions in Design, the Symposium on Design Research and Semiotics*. UIAH, Helsinki.
130. [Hammer91] Hammer, N., Lengyel, S. (1991). Identifying Semantics Markers in Design Products: The Use of Eye-Movement recording in Industrial Design. *Oculomotor Control and Cognitive Processes*. Elsevier Science Publishers B.V. (North-Holland)
131. [Hammer92] Hammer, N. (1992). Möglichkeiten und Grenzen der Überprüfung von Designprodukten durch Okulometrie. *Die Blaue Eule*, Essen.
132. [Hawkins90] Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., and Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of experimental psychology : Human perception and performance*, 16, 802-811.
133. [Hawkins90b] Hawkins, S.A., Hastie, R. (1990). Hindsight: biased judgements of past events after the outcomes are known, *psychological bulletin*, 107 (may), 311-327.
134. [Hayhoe00] Hayhoe, M. M. (2000). Vision using routines : A functional account of vision. *Visual Cognition*, 7, 43-64
135. [Hayhoe03] Hayhoe, M.M., Shrivastava, A., Mruczek, R., & Pelz, J. (2003). Visual memory and motor planning in a natural task. *Journal of vision*, 3, 49-63
136. [Hayhoe05] Hayhoe, M. M., & Ballard, D. H. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 188-194.
137. [Heaps99] Heaps, C., & Handel, C.H. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 299-320.
138. [Hebb68] Hebb, D. O. (1968). Concerning imagery. *Psychological Review*, 75, 466-477

139. [Hekkert03] Hekkert, P., Snelders, H.M.J.J., & van Wieringen, P.C.W. (2003). Most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference. *British Journal of Psychology*, 94, 111-124.
140. [Hekkert06] Hekkert P. (2006). Design aesthetics : principles of pleasure in design (in press).
141. [Hekkert95] Hekkert, P. & Snelders, D. (1995). Prototypicality as an explanatory concept in aesthetics: A reply to Boselie (1991). *Empirical Studies of the Arts*, 13, 149-160
142. [Hekkert96] Hekkert, P. & Wieringen, P.C.W., van. (1996). Beauty in the eye of expert and nonexpert beholders: A study in the appraisal of art. *American Journal of Psychology*, 109, 389-407.
143. [Helmholtz24] Helmholtz, H. (1924a). Helmholtz's treatise on physiological optics. Vol. 1. (Trans. J. P. C. Southall) New York : Optical Society of America.
144. [Henderson03] Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, 7, 498-504.
145. [Henderson03b] Henderson, J. M., Williams, C. C., & Castelano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception & Psychophysics*, 65, 725-734
146. [Henderson03c] Henderson, J. M., Hollingworth, A. (2003). Eye Movements, visual memory, and scene representation. In M. S. Peterson & G. Rhodes (Eds.). *Analytic and holistic processes in the perception of faces, objects and scenes* (pp. 356-377). New York: Jal/Ablex
147. [Henderson04a] Henderson, J. M., and Ferreira, F. (2004a). Scene perception for psycholinguists. In J. M. Henderson, and F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
148. [Henderson06] Henderson, J. M. Visual Saliency Does Not Account for Eye Movements during Visual Search in Real-World Scenes ; Chapter to appear in: Roger van Gompel, Martin Fischer, Wayne Murray, and Robin Hill (Eds). *Eye Movement Research: Insights into Mind and Brain*. Elsevier.
149. [Henderson92] Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 260-283). New York Springer-Verlag
150. [Henderson97] Henderson, J.M., Hollingworth A. (1997). Eye movement during scene viewing : An overview. Technical Report, 5. Michigan State University Eye Movement Laboratory
151. [Henderson98] Henderson, J. M., Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 269-283). Oxford: Elsevier.
152. [Henderson99] Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243-271.
153. [Henderson99b] Henderson, J. M., Weeks, P. A. Jr., & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 210-228.
154. [Hess90] Hess, T.M., Slaughter, J.S. (1990). Schematic knowledge influences on memory for scene information in young and older adults. *Developmental psychology*, 26, 855-865.
155. [Hoch86] Hoch, S.J., Ha, Y.W. (1986). Consumer learning : advertising and the ambiguity of product experience. *Journal of consumer research*, 13, 221-223
156. [Hoffman95] Hoffman, J. E., & Subramaniam, B. (1995). The role of visual-attention in saccadic eye-movements. *Perception & Psychophysics*, 57(6), 787-795.
157. [Holbrook82] Holbrook, M. B., & Hirschman, E. C. (1982). The Experiential Aspects of Consumption: Consumer Fantasies, Feelings and Fun. *Journal of Consumer Research*, 9, 132-140
158. [Hollingworth01a] Hollingworth, A. (2001) Change detection in the flicker paradigm: the role of fixation position within the scene. *Mem.Cognit.* 29, 296-304
159. [Hollingworth01b] Hollingworth, A. (2001) To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. *Psychon. Bull. Rev.* 8, 761-768
160. [Hollingworth02] Hollingworth, A. Henderson, J.M. (2002) Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 113-136
161. [Hollingworth06] Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning Memory and Cognition*, 32, 58-69.
162. [Hollingworth98] Hollingworth, A., Henderson, J. (1998). Does consistent scene context facilitate object perception ? *Journal of experimental psychology : General*, 127, 398-415.
163. [Hollingworth99] Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica*, 102, 319-343.
164. [Hooge00] Hooge I.T.C., Frens M.A. (2000). Inhibition of saccade return (ISR): spatio-temporal properties of saccade programming. *Vision Research*, Volume 40, Number 24, October 2000 , pp. 3415-3426(12).
165. [Hornof02] Hornof, A., Halverson, T. (2002) Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior research methods, instruments & computers*. 2002, vol. 34, no4, pp. 592-604

166. [Humphrey89] Humphrey, G.W., & Bruce, V. (1989). *Visual cognition : computational, experimental, and neuropsychological perspectives*. Lawrence Erlbaum Associates Ltd., Hove, UK.
167. [Humphreys06] Humphreys, K., Gosselin, F., Schyns, P. G., Johnson, M. H. (2006). Using "Bubbles" with babies: A new technique for investigating the informational basis of infant perception. *Infant Behavior & Development* 29 (2006) 471–475.
168. [Hutchinson86] Hutchinson, J. W. (1986). Discrete attribute models of brand switching. *Marketing Science*, 5(4), 350–371.
169. [Hyrskykar03] Hyrskykari, A., Majaranta, P., Rähkä, K. –J. (2003). Proactive response to eye movements, *Proceedings of Interact 2003*, Zurich, pp. 129-136, September 2003
170. [Hyvärinen01] Hyvärinen, A., Karhunen, J., Oja, E. (2001) *Independent Component Analysis*; 2001 John Wiley & Sons.
171. [Ijsselsteijn01] Ijsselsteijn, W.A., de Ridder, H., Freeman, J., Avons, S., & Bouwhuis, D. (2001). Effects of stereoscopic presentation, image motion and screen size on subjective and objective corroborative measures of presence. *Presence: Teleoperators and Virtual Environments* 10, 298-311.
172. [Ilg93] Ilg, U. J., Hoffmann, K. P. (1993). Motion perception during saccades. *Vision research*, 33, 211-220.
173. [Intraub81] Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 604-610.
174. [Intraub89] Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 179–187.
175. [Intraub93] Intraub, H., & Bodamer, J.L. (1993). Boundary extension : Fundamental aspect of pictorial representation or encoding artefact ? *Journal of experimental psychology: Learning, Memory, and cognition*, 19, 1387-1397
176. [Intraub96] Intraub, H., Gottesman, C.V., Willey, E.V. & Zuk, I.J. (1996). Boundary extension for briefly glimpsed photographs: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory & Language*, 35(2), 118-134
177. [Intraub98] Intraub, H., Gottesman, C.V., & Bills, A.J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of experimental psychology: Learning, Memory, and cognition*, 24, 307-317
178. [Irwin02] Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics*, 64, 882–895.
179. [Irwin04] Irwin, D. E. (2004). Fixation Location and Fixation Duration as Indices of Cognitive Processing. In J. M. Henderson, & F. Ferreira (Eds). *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
180. [Irwin92] Irwin, D. E. (1992). Memory for position and identity across eye-movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 307–317.
181. [Irwin96] Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.
182. [Itti00] Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.
183. [Itti01] Itti, L., Koch, C. (2001) *Computational Modeling of Visual Attention*, *Nature Reviews Neuroscience*, Vol. 2, No. 3, pp. 194-203, Mar 2001.
184. [Itti05] Itti, L., Rees, G., Tsotsos, J. K. (2005) *Neurobiology of Attention*, pp. 1-744, San Diego, CA: Elsevier, Jan. 2005.
185. [Itti06] Itti, L. (2006) Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, in press.
186. [Itti06b] Itti, L., Baldi, P. (2006) Bayesian Surprise Attracts Human Attention, In: *Advances in Neural Information Processing Systems*, Vol. 19 (NIPS*2005), pp. 1-8, Cambridge, MA:MIT Press, 2006.
187. [Jaccard86] Jaccard, J. Brinberg, D., Ackerman L. J. (1986). Assessing Attribute Importance: A Comparison of Six Methods ; *The Journal of Consumer Research*, Vol. 12, No. 4 (Mar., 1986), pp. 463-468
188. [Jacob03] Jacob, R. J. K., Karn, K. S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), in *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*, J. Hyona, R. Radach, H. Deubel (Eds.), Oxford, Elsevier Science
189. [Jacob95] Jacob, R.J.J. (1995). eye tracking in advanced interface design. In Barfield, W., & Furness, T. (Eds.), *advanced interface design and virtual environments*, pp.258-288. Oxford University Press, Oxford
190. [Jacobson79] Jacobson, J. Z., & Dodwell, P. C. (1979). Saccadic eye movements during reading. *Brain and Language*, 8(3), 303-314.
191. [Jacoby89] Jacoby, L. L., Kelley, C. Dywan, J. (1989). Memory attributions. In H.L. Roediger & F.I.M. Craik (Eds.) *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.
192. [Jacoby89b] Jacoby, L.L., Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126–135.

193. [Jaimes00] Jaimes, A., Chang S-F. (2000) A Conceptual Framework for Indexing Visual Information at Multiple Levels, IS&T/SPIE Internet Imaging, Vol. 3964, San Jose, Ca, 2000
194. [James90] James, W. (1890) Principles of Psychology, Dover Publications
195. [Jorgensen96] Jørgensen, C. (1996) Indexing Images: Testing an Image Description Template. ASIS 1996 Annual Conference Proceedings, October 19-24, 1996.
196. [Julesz86] Julesz, B. (1986). Stereoscopic vision, *Vision Research* 26(9), pp. 1601-1612. Twenty-fifth Anniversary Issue.
197. [Just76] Just, M. A. and Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8:441–480.
198. [Kahneman73] Kahneman, D. (1973). Attention and effort. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
199. [Kanwisher01] Kanwisher, N. (2001). Faces and places: Of central (and peripheral) interest. *Nature Neuroscience*, 4, 455–456.
200. [Kapferer85] Kapferer J. N. et Laurent G. (1985), Consumers' Involvement Profile. Empirical Results, *Advances in Consumer Research*, 290-295.
201. [Kardes02] Kardes, F.R., Sanbonmatsu, D.M. Cronley, M.L., Houghton, D.C. (2002). Consideration Set Overvaluation: When Impossibly Favorable Ratings of a Set of Brands are observed. *Journal of Consumer Psychology*, 12 (4), 353–61.
202. [Kaufman69] Kaufman, L. & Richards, W. (1969). Spontaneous fixation tendencies for visual forms. *Perception and Psychophysics*, 5, 85-88
203. [Keinonen98] Keinonen, T (1998): One-dimensional usability – One-dimensional usability – Influence of usability on consumer's product preference : University of Art and Design Helsinki, UIAH A21.
204. [Kelly84] Kelly, D. (1984). Retinal inhomogeneity: I. spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A*, 74(1):107–113.
205. [Keogh02] Keogh, E. (2002). Exact indexing of dynamic time warping. August 2002, Proceedings of the 28th international conference on Very Large Data Bases - Volume 28 VLDB '2002
206. [King94] King, M.F., Balasubramanian, S.K. (1994). The Effects of Expertise, End Goal, and Product Type on Adoption of Preference Formation Strategy. *Journal of the Academy of Marketing Science*, Vol. 22, No. 2, 146-159 (1994)
207. [Klinger94] Klinger, M. R., Greenwald A. G. (1994). Preferences need no inferences?: 'The cognitive basis of unconscious mere exposure effects. In P.M. Niedenthal & S. Kitaynma (Eds.); *The heart's eye* (pp67-85) San Diego: Academic Press.
208. [Koch85] Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry, *Human Neurobiology*, 4:219-227.
209. [Koenderink99] Koenderink, J. J. and van Doorn, A. (1999). The structure of locally orderless images. *International Journal of Computer Vision*, 31(2/3):159–168.
210. [Kohonen97] Kohonen, T.(1997). Self-Organizing Maps, New York : Springer-Verlag, 1997.
211. [Koivunen04] Koivunen, K., Kukkonen, S., Lahtinen, S., Rantala, H., Sharmin, S. (2004). Towards Deeper Understanding of How People Perceive Design in Products. CADE2004 Web Proceedings of Computers in Art and Design Education Conference, 2004
212. [Kolb01] Kolb, H., Fernandez, E., Nelson, R. (2001) *WebVision - The Organization of the Retina and Visual System*
213. [Komatsu92] Komatsu, L.K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, 112(3), 500-526.
214. [Kowler06] Kowler, E. (2006). Attention and Eye Movements. In *New encyclopedia of neuroscience* (volume editor R. Krauzlis), Elsevier.
215. [Kowler95] Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35, 1897-1916
216. [Kullback59] Kullback, S.(1959). *Information Theory and Statistics*, John Wiley Publications, 1959.
217. [Kunsthilson80] Kunsthilson, W. R. Zajonc, R. B. (1980). Affective Discrimination of Stimuli That Cannot Be Recognized. *Science* 207, 557-558
218. [Laeng02] Laeng, B., & Teodorescu, D.S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, 26, 207-231.
219. [Lancaster66] Lancaster, Kelvin J. (1966). A New Approach to Consumer Theory .*The Journal of Political Economy*, Vol. 74, No. 2 (Apr., 1966), pp. 132-157
220. [Land01] Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, 11(15), 1215-1220
221. [Land01b] Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25-26), 3559–3565.
222. [Land94] Land, M. F., Lee, D. N. (1994). Where we look when we steer. *Nature*, 369, 742-744.

223. [Langlois90] Langlois, J. H. ; Roggman, L.A. (1990). Attractive Faces Are Only Average. *Psychological Science* 1(2): 115-121.
224. [Lariani00] Lariano, S. (2000) Perception et interprétation de sections et blocs sismiques : oculométrie et analyse d'images. Thèse de doctorat Université Joseph Fourier, LIS/INPG, Grenoble.
225. [Laudati00] Laudati, P. (2000) Perception de l'image de la ville et qualifications sémantique des espaces. Thèse de doctorat, Université de Valenciennes et du Hainaut Cambresis.
226. [Law 04] Law, B., Atkins, M. S., Kirkpatrick, A. E., Lomax, A. J., AND Mackenzie, C. L. (2004). Eye Gaze Patterns. Differentiate Novice and Experts in a Virtual Laparoscopic Surgery Training Environment. In *Eye Tracking. Research & Applications Symposium*. ACM, 41–47.
227. [Lécuyer01] Lécuyer, A., Burkhardt, J.M., Coquillart, S., Coiffet, P. (2001). "Boundary of Illusion" : an Experiment of Sensory Integration with a Pseudo-Haptic System, *IEEE Int. Conf. on Virtual Reality*, pages 115-122, Yokohama, Japan, 2001.
228. [Leder05] Leder, H. & Carbon, C. C. (2005). Dimensions in appreciation of car interior design. *Applied Cognitive Psychology*, 19(5), 603-618.
229. [Lee00] Lee & Yu (2000). An information-theoretic framework for understanding saccadic behaviors. *Advanced in Neural Processing Systems*, 12, 834-840.
230. [Lee01] Lee, A. Y. (2001). The mere exposure effect: an uncertainty reduction explanation revisited. *Personality and Social Psychology Bulletin*, 27(10), 1255–1266.
231. [Leek03] Leek, E.C. et al. (2003) Inhibition of return for objects and locations in static displays. *Percept. Psychophys.* 65, 388–395
232. [Lefkoff93] Lefkoff-Hagius R., Mason, C.H. (1993).« Characteristics, beneficial, and image attributes in consumer judgements of similarity and preference », *Journal of Consumer Research*, 20, 100-110
233. [LeNy79] Le Ny, J.F. (1979). *La sémantique psychologique*. Paris, Presses Universitaires de France.
234. [Levenshtein66] Levenshtein, V. I.(1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10 (1966):707–710
235. [Levine97] Levine, L.J. (1997). Reconstructing memory for emotions. *Journal of experimental psychology : General*, 126(june),165-177.
236. [Li02] Li, F-F. et al. (2002) Natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9596–9601
237. [Liu01] Liu, A., Salvucci, D. (2001). Modeling and prediction of human driver behavior. 9th. Intl. conference on human-computer interaction. New Orleans, LA., Aug. 2001
238. [Liversedge00] Liversedge, S.P. & Findlay, J.M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6-14
239. [Li-Z02] Li, Z. (2002) A saliency map in primary visual cortex. *Trends Cogn. Sci.* 6, 9–16
240. [Locher01] Locher, P., Cornelis, E., Wagemans, J., & Stappers, P. J. (2001). An empirical investigation of the role of balance in the creation of visual designs by adults. *Empirical Studies of the Arts*, 19, 213-227.
241. [Locher03] Locher, P. J. (2003). An empirical investigation of the visual rightness theory of picture perception. *Acta Psychologica*, 114(2), 147-164.
242. [Loftus78] Loftus, G.R. ; Mackworth, N.H.. Cognitive determinants of fixation location during picture viewing. *Journal of experimental psychologie : human perception and performance*, 4(4) : 565-572, 1978
243. [Loftus81] Loftus, G. R. (1981). Tachistosopic simulations of eye fixations on pictures. *Journal of Experimental Psychology*, 7(5), 369-376
244. [Loftus83] Loftus, G.R., Nelson, W.W., & Kallman, H.J. (1983). Differential acquisition rates for different types of information from pictures. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 35, 187-198.
245. [Lynch06] Lynch, P.J. (2006). *Medical illustrator*. Creative Commons Attribution 2.5 License 2006. <http://www.wikipedia.org>
246. [Mace06] Macé, Marc. (2006). Représentations visuelles précoces dans la catégorisation rapide de scènes naturelles chez l'homme et le singe. Thèse de l'Université Paul Sabatier Toulouse.
247. [MacKenzie86] MacKenzie, S.B. (1986). "The Role of Attention in Mediating the Effect of Advertising on Attribute Importance," *Journal of Consumer Research: An Interdisciplinary Quarterly*, University of Chicago Press, vol. 13(2), pages 174-95, September.
248. [Mackworth67] Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, 2, 547-552.
249. [Majaranta02] Majaranta, P., Raiha, K.-J. (2002). Twenty Years of Eye Typing: Systems and Design Issues. In *Eye Tracking Research & Applications: Porceedings of the symposium ; on ETRA 2002* (pp. 15-22). New York: ACM Press

250. [Mandler76] Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology*, 2, 38–48.
251. [Mannan95] Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, 9, 363-386.
252. [Mannan96] Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, 10, 165-188.
253. [Mannan97] Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, 26, 1059-1072.
254. [Marr82] Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman
255. [Martin40] Martin, G. (1940). Eye movements as the cause of faulty steering of airplanes, automobiles, and bicycles. *Graefes Archiv fuer Ophthalmologie*
256. [Martindale88] Martindale, C., Moore, K., (1988). Priming, prototypicality, and preference. *Journal of experimental psychology : human perception and performance*, 14, 661-670.
257. [Martindale90] Martindale, C., Moore, K., & Borkum, J. (1990). Aesthetic preference: anomalous findings for Berlyne's psychobiological theory. *American Journal of Psychology*, 103(1), 53–80.
258. [Masson98] Masson, M. E. J., Caldwell, J. I. (1998). Conceptually driven encoding episodes create perceptual misattributions. *Acta Psychologica*, 98, 183-210.
259. [Matin74] Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81, 899-917.
260. [Mazer03] Mazer, J.A. and Gallant, J.L. (2003) Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron* 40, 1241–1250
261. [McConkie02] McConkie & Loschky, (2002), Perception onset time during fixation in free viewing. *Behavior research methods, instruments & computer*, 34, 481-490
262. [McConkie96] McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 563–581.
263. [McCotter05] McCotter, Gosselin, F., Sowden, P., and Schyns, P.G. (2005) The use of visual information in natural scenes, *Vis. Cogn.*, 12: 938-953.
264. [McPeck00] McPeck, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, 40, 2499–2516
265. [Megaw79] Megaw, E. D., Richardson, J. (1979). Eye movements and industrial inspection. *Applied Ergonomics* 10, 3, 145–154.
266. [Melcher01] Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, 41(25–26), 3597–3611.
267. [Melcher99] Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research*, 39, 2929–2946.
268. [Menon01] Menon, G., Raghubir, P. (2001). When Automatic accessibility meets conscious content : implications for the accessibility-diagnostics framework. New York University working paper
269. [Merviel97] Merviel, S., (1997), *La conception en communication, Méthodologie qualité*, éd. Hermes, Paris, 1997, p.80
270. [Mesulam90] Mesulam, M.M. (1999) Spatial attention and neglect, parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354, 1325–1346
271. [Miller60] Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart & Winston.
272. [Molnar68] Molnar, F. (1968). Towards a science in art. In A. Hill (Ed.), *Data: directions in art, theory of aesthetics*. An anthology (pp. 204-213). London: Faber & Faber.
273. [Moore04] Moore, T. and Fallah, M. (2004) Microstimulation of the frontal eye field and its effects on covert spatial attention. *J. Neurophysiol.* 91, 152–162
274. [Morrison81] Morrison, R. E., & Inhoff, A. W. (1981). Visual factors and eye movements in reading. *Visible Language*, 15(2), 129-146
275. [Morrone05] Morrone, M. C., Ross, J., Burr, D. (2005) Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience* vl.8, nb. 7; 19/06/2005
276. [Muller05] Muller, N., G., Mollenhauser, M., Rosler, A., Kleinschmidt, A. (2004). The attentional field has a Mexican hat distribution. *Vision research* 45 (2005) 1129-1137.
277. [Munoz00] Munoz, D.P. et al. (2000) On your mark, get set: brainstem circuitry underlying saccadic initiation. *Can. J. Physiol. Pharmacol.* 8, 934–944
278. [Murphy93] Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality & Social Psychology*, 64(5), 723-739.

279. [Najemnik05] Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387–391.
280. [Navalpakkam05] Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2), 205–231.
281. [Nelson80] Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 391–399.
282. [Nodine04] Nodine, C. F., & Krupinski, E. A. (2004). How do viewers look at artworks? *Bulletin of Psychology and the Arts*, 4, 65-68.
283. [Noton71a] Noton, D., & Stark, L. W. (1971a). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision research*, 11, 929–942
284. [Noton71b] Noton, D., & Stark, L. W. (1971b). Scanpaths in eye movements during perception. *Science*, 171, 308–311.
285. [O'Regan00] O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, 7(1–3), 191–211.
286. [O'Regan83] O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does transsaccadic fusion exist. *Vision Research*, 23(8), 765–768.
287. [Oliva00] Oliva, A., & Schyns, P. G. (2000). Colored diagnostic blobs mediate scene recognition. *Cognitive Psychology*, 41,176-210.
288. [Oliva01] Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal in Computer Vision*, 42, 145-175.
289. [Oliva03] Oliva, A., & Torralba, A., Castelhana, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing*, pp. 253–256
290. [Oliva04] Oliva, A. (2004). Gist of the scene. In L .Itti, G . Rees & J.K. Tsotsos (Eds.), *Neurobiology of attention*, San diego, CA; Elsevier
291. [Oliva04b] Oliva, A., Mack, M.L., Shrestha, M., & Peeper, A. (2004) Identifying the perceptual dimensions of visual complexity in scenes. *Proc. of the 26th Annual Meeting of the Cogn. Sci. Soc.*, Chicago, August.
292. [Oliva06] Oliva, A. & Torralba, A. (2006). Building the Gist of a Scene: The Role of Global Image Features in Recognition. *Progress in Brain Research: Visual perception*, 155, 23-36.
293. [Oliva97] Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, 34, 72-107.
294. [Oliva99] Oliva A., Torralba A., Guérin-Dugué A. et Héroult J. (1999) Super-Ordinate Representation of Scenes from Power Spectrum Shapes. *CIR'99, 2nd Conf. On image retrieval Newcastle, February 25-29, 1999.*
295. [O'Regan92] O'Regan, J. K. (1992). Solving the “Real” Mysteries of Visual Perception: The World as an Outside Memory. *Canadian Journal of Psychology*, 46, 461-488.
296. [Pannasch07] Pannasch, S., Helmert, J.R. & Velichkovsky, B.M. (2007). The relationship of spatio-temporal characteristics of eye movements in scene perception. *Perception*, 36, (suppl), 29.
297. [Park94] Park C. W., Mothersbaugh, D. L. et Feick L. (1994), *Consumer Knowledge Assessment*, *Journal of Consumer Research*, 21 (June), 71-82
298. [Parker78] Parker, R. E. (1978). Picture processing [during recognition]. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 284-293.
299. [Parkes01] Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *J. Nat. Neurosci.* 4, 739–744
300. [Parkhurst02] Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107-123.).
301. [Parkhurst03] Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, 6, 125-154.).
302. [Parzen62] Parzen E. (1962). On estimation of a probability density function and mode, *Ann. Math. Stat.* 33, pp. 1065-1076.
303. [Pavlov27] Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Translated and Edited by G. V. Anrep. London: Oxford University Press.
304. [Payne93] Payne, J.W., Bettman, J.R., Johnson, J.E. (1993). *The adaptative decision making*. Cambridge University Press, 1993.
305. [Peter93] Peter, J. P., Olson, J. C. (1993). *Consumer Behavior and Marketing Strategy*, 3rd, IRWIN, Homewood, Illinois, pp.474-482.
306. [Peters07] Peters, R. J., Itti, L. (2007) Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception*, Vol. in press, No. preprint, May 2007, Pages 1–21.

307. [Peterson01] Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics*, 63, 1239-1249
308. [Petty93] Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology*, 29, 137-165
309. [Pierce1907] Pierce, C.S. (1907) Pragmatism) v.1907 - MS 318 - source : <ftp://ftp.univ-perp.fr/pub/semiotics/marty/76-fr.zip>
310. [Pollatsek84] Pollatsek, A., Rayner, K. and Collins, W. E. (1984). Integrating pictorial information across saccadic eye movements. *J. Exp. Psych., Gen.*, 113:426-442.
311. [Pollatsek92] Pollatsek, A., & Rayner, K. (1992). What is integrated across fixations? In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 166–191). New York: Springer Verlag.
312. [Posavac05] Posavac, S. S., Kardes, F.R., Sanbonmatsu, D.M., Fitzsimons G.J. (2005). Blissful Insularity: When Brands are Judged in Isolation from Competitors ; *Marketing Letters*, 16 (April), 87-97.
313. [Posner80] Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psych.*, 32:3-25
314. [Posner84] Posner, M. I., Walker, J. A., Friedrich, F. J. and Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention. *J. Neurosci.*, 4:1863-1874
315. [Posner85] Posner, M.I. et al. (1985) Inhibition of return: neural basis and function. *Cogn. Neuropsychol.* 2, 211–228
316. [Potter75] Potter, M.C. (1975). Meaning in visual search. *Science*, 187(4180), 965-966.
317. [Potter76] Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.
318. [Qvarfordt05] Qvarfordt P., Zhai. S., *Conversing with the User Based on Eye-Gaze Patterns*, CHI 2005, april 2-7, 2005, Portland, Oregon, USA
319. [Raidt07] Raidt, S., G. Bailly and F. Elisei (2007). Mutual gaze during face-to-face interaction. *Auditory-visual Speech Processing*, Hilvarenbeek, The Netherlands
320. [Raj05] R., Geisler, W. S., Frazor, R. A., & Bovik, A. C.(2005). Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 22, 2039–2049.
321. [Rajashekara04] Rajashekara, U., Cormackb, L. K., Bovik, A. C. (2004). Point of Gaze Analysis Reveals Visual Search Strategies
322. [Ramachandran99] Ramachandran, V.S., Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *Journal of consciousness studies*, 6, 15-51.
323. [Rao02] Rao, R. P.N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, 42(1), 1447-1463.
324. [Rasche02] Rasche, C., Koch, C. (2002). Recognizing the gist of a visual scene : possible perceptual and neural mechanisms. *Neurocomputing*, 44-46, 979-984.
325. [Rayner78] Rayner, K. (1978). Foveal and parafoveal cues in reading. In J. Requin (Ed.). *Attention and performance* (Vol. 7, pp. 149-162). Hillsdale, NJ : Lawrence Erlbaum Associates, Inc.
326. [Rayner83] Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades. *Perception & Psychophysics*, 34(1), 39–48.
327. [Rayner98] Rayner, K. (1998). Eye movements in reading and information processing. 20 years of research. *Psychol. Bull.*, 124:372-422.
328. [Reber98] Reber, R., Winkelman, P. Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9, 45-48.
329. [Reinagel99] Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, 10, 341–350.
330. [Reingold02] Reingold, E. M., Stampe, D. M. (2002). Saccadic inhibition in voluntary and reflexive saccades. *Journal of Cognitive neuroscience*, 2002; 14; 371-388
331. [Renninger03] Renninger, L. W. (2003). Parts, objects and scenes: computational models and psychophysics; PhD thesis; university of California, Berkeley.
332. [Renninger04] Renninger, L. W. (2004). When is scene identification just texture recognition? *Vision Research* 44 (2004) 2301-2311.
333. [Renninger05a] Renninger, L. W., Verghese, P., & Coughlan, J. (2005a). Eye movements can be understood within an information theoretic framework. *Computational & Systems Neuroscience (Cosyne05)*, Salt Lake City, UT.
334. [Renninger05b] Renninger, L.W., Coughlan, J.M., Verghese, P. & Malik, J. (2005). An information maximization model of eye movements. *Advances in Neural Information Processing Systems*, 17, 1121-1128.
335. [Renninger07] Renninger, L. W., Verghese, P., Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of vision* 7(3) :6 ;1-17.
336. [Rensink00] Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7, 17–42.

337. [Rensink00a] Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition*, 7(1–3), 17–42.
338. [Rensink00b] Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7(1–3), 127–145.
339. [Rensink00c] Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research*, 40(10–12), 1469–1487.
340. [Rensink97] Rensink, R., O'Regan, J. K. and Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.*, 8:368-373
341. [Repp97] Repp, B.H. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, 14, 419-444.
342. [Rhodes96] Rhodes, G., Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science*, 7, 2, 105-110.
343. [Robinson85] Robinson, L. B., Hastie, R. (1985). Revision of Beliefs when a Hypothesis Is Eliminated from Consideration ; *Journal of Experimental Psychology: Human Perception and Performance*, 11 (August), 443–56.
344. [Rojas05] Rojas, J., Massot, C., Herault, J. (2005) Interactions corticales : des illusions visuelles aux scènes naturelles. Mémoire de stage de master information, cognition et apprentissages spécialité sciences cognitives, Laboratoire des Images et Signaux, INP Grenoble.
345. [Rosander04] Rosander, K., VonHofsten, C. (2004). Infant's emerging ability to represent occluded object motion. *Cognition*, 91, 1-22.
346. [Rosch73] Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, 4, 328-350.
347. [Rosch75] Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology*, 104, 192-233.
348. [Rosch78] Rosch, E. (1978). Principles of categorization . In E., Rosch, B., Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.
349. [Rousselet02] Rousselet, G. A., M. Fabre-Thorpe and S. J. Thorpe (2002). Parallel processing in high-level categorization of natural images. *Nat Neurosci* 5(7): 629-30.; Kirchner, H. and S. J. Thorpe (2006). Ultra-rapid classification and saccadic eye movements: Visual processing speed revisited. *Vision Res* 46(11): 1762-76
350. [Rovamo79] Rovamo, J. & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37, 495-510.
351. [Rucci07] Rucci, M., Iovin, R., oletti, M., Santini, F. (2007), Miniature Eye Movements Enhance Fine Spatial Detail, *Nature*. 447(7146), 851-854.
352. [Russell91] Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin* 110,426-450.
353. [Russo00] Russo, J.E., Meloy, M.G., Wilks, J.T. (2000). Predecisional distortion of information by auditors and salespersons. *Management science*, 46(1), 13-27.
354. [Russo98] Russo, J.E., Meloy, M.G., Medvec, V.H. (1998). The distortion of product information during brand choice. *Journal of marketing research*, 35(4), 438-452.
355. [Sadler73] Sadler, O., Tesser, A. (1973). Some Effects of Saliency and Time upon Interpersonal Hostility and Attraction ; *Sociometry*, 36 (March), 99–112.
356. [Salojarvi03] Salojarvi, J., Kojo, I., Simola, J., Kaski, S. (2003). Can relevance be inferred from eye movements in information retrieval ?, in *Proceedings of the Workshop on Self-Organizing Maps (WSOM'03)*, Hibikino, Kitakyushu, Japan, September 2003, pp. 261–266
357. [Salvucci00] Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.
358. [Salvucci01] Salvucci, D., Anderson, J. (2001). Automated eye-movement protocol analysis. *Human-Computer Interaction* 16, 39–86.
359. [Sanbonmatsu98] Sanbonmatsu, D.M., Posavac, S.S., Kardes, F.R., Mantel, S.P. (1998). Selective Hypothesis Testing ; *Psychonomic Bulletin and Review*, 5 (June), 197–220.
360. [Santella02] Santella, A., DeCarlo, D. (2002). Abstracted painterly renderings using eye-tracking data. *Second International Symposium on Non-Photorealistic Animation and Rendering (NPAR)*
361. [Santella04] Santella, A., DeCarlo, D. (2004). Robust clustering of Eye Movement recordings for quantification of visual interest. *Eye tracking research and application 2004 – archives*
362. [Schachter64] Schachter, S. (1964). The Interaction Of Cognitive And Physiological Determinants Of Emotional State. *Advances In Experimental Social Psychology* 1(1): 49-80
363. [Schall97] Schall, J.D. (1997) Visuomotor areas of the frontal lobe. In *Extrastriate Cortex of Primates, Cerebral Cortex* (Rockland, K.S. et al., eds), pp. 527–538, Plenum Press
364. [Schoonard73] Schoonard, J. W., Gould, J. D., & Miller, L.A. (1973). Studies of visual inspection, *Ergonomics*, 16(4), 365-379

365. [Schwarz90] Schwarz, N. (1990). Feeling as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition* (pp. 527-561). New York: Guilford Press.
366. [Schyns94] Schyns, P.G. & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4), 195-200.
367. [Schyns97] Schyns, P. G. & Rodet, (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23, 1-16.
368. [Schyns98] Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, 21, 1-53.
369. [Schyns99] Schyns, P. G. & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69, 243-265.
370. [Sereno06] Sereno, S. C. & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7, 489-493.)
371. [Shannon48] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423 and 623-656.
372. [Shannon49] Shannon, C. E. (1949). *Communication Theory of Secrecy Systems*, *Bell System Technical Journal*, Vol 28, pp. 656-715, Oct 1949.
373. [Sharmin04] Sharmin, S. (2004). *Studies of Human Perception on Design Products*, university of tampere, December 2004.
374. [Shepherd86] Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye-movements and spatial attention. *Quarterly Journal of Experimental Psychology*, 38A, 475-491.
375. [Shimojo03] Shimojo, S. & Simion, C., Shimojo, E., Scheier, S. (2003). Gaze bias both reflects and influences preference - *Nature Neuroscience*, Vol. 6(12), Dec. 2003, pp. 1317-1322.
376. [Simion05] Simion, C. (2005) *Orienting and preference: an enquiry into the mechanisms underlying human decision making*. Phd Thesis, Caltch library,
377. [Simons96] Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, 7(5), 301-305.
378. [Simonson00] Simonson, I., Nowlis, S.M. (2000). The Role of Explanations and Need for Uniqueness in Consumer Decision Making: Unconventional Choices Based on Reasons ; *Journal of Consumer Research*, 27 (June), 49-68.
379. [Simonson89] Simonson, I. (1989). Choice based on reasons : the case of attraction and compromise effects. *Journal of consumer research*, 16 (september), 158-174
380. [Smith98] Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 269-322). Boston: McGraw-Hill.
381. [Sokolov63] Sokolov, E. N. (1963). "Higher Nervous Functions - Orienting Reflex." *Annual Review Of Physiology* 25:
382. [Solomon94] Solomon, M.R., Schiffman, L.G., Kanuk, L.L. (1994) *Consumer Behavior*, 1994 U.S.A.
383. [Sparks02] Sparks, D.L. (2002) The brainstem control of eye movements. *Nat. Rev. Neurosci.* 3, 952-964
384. [Specht41] Specht, H. (1941). Fatigue and hours of service of interstate truck drivers. V: Eye movements and related phenomena. *Public Health Bulletin* Washington
385. [Stampe95] Stampe, D. M., & Reingold, E. M. (1995). Selection by looking: A novel computer interface and its application to psychological research. In J. M. Findlay, R. Walker & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 467-478). Amsterdam: Elsevier
386. [Stang74] Stang, D. J. (1974). Methodological factors in mere exposure research. *Psychological Bulletin*, 81(12), 1014-1025
387. [Stern93] Stern, J. A. (1993). The eyes : Reflector of attentional processes. *CSERIAC Gateway*, IV(4), 7-12. A Synopsis by J. Skelly
388. [Strahan01] Strahan, E.A., Spencer, S.J., Zanna, M.P. (2001). Subliminal priming and persuasion : striking while the iron is hot. University of Waterloo, Psychology department working paper.
389. [Svenson92] Svenson, O., Benthorn, L.J. (1992). Consolidation process in decision making : post-decision changes in attractiveness of alternatives. *Journal of economic psychology*, 13 (june), 315-327
390. [Tatler03] Tatler, B. W., Gilchrist, I. D., Rusted, J. (2003). The time course of abstract visual representation. *Perception*, 32(5), 579-592.
391. [Tatler05] Tatler, B. W., Baddeley, R. J., Gilchrist, I. D. (2005). Visual correlates of fixation selection : effects of scale and time. *Vision research* 45 (2005) 643-659.
392. [Tatler05b] Tatler, B. W., Gilchrist, I. D., Land, M. F. (2005). Visual memory for objects in natural scenes : From fixations to object files. *The Quarterly journal Of Experimental Psychology* ; 2005, 58A (5), 931-960
393. [Tatler06] Tatler, N. W., Baddeley, R. J., Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision research* (2006). In press.

394. [Thiele02] Thiele, A., Henning, M., Buischik, K., Hoffman, P. (2002). Neural mechanisms of saccadic suppression. *Science*, 295, 2460-2462.
395. [Thompson05] Thompson, K.G. and Bichot, N.P. (2005) A visual saliency map in the primate frontal eye field. *Prog. Brain Res.* 147, 251
396. [Thorpe 97] Thorpe S.J., Gautrais J. (1997). Rapid Visual Processing using Spike Asynchrony. In *Neural Information Processing Systems*, ed. M.C. Mozer, M.I. Jordan, T. Petsche, pp. 901-7. Cambridge: M.I.T. Press
397. [Thorpe96] Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522.
398. [Tobacyk79] Tobacyk, J., Bailey, L., & Myers, H. (1979). Preference for paintings and personality traits. *Psychological Reports*, 45, 787-793.
399. [Torralba01] Torralba, A., Sinha, P. (2001). Statistical context priming for object detection. Proceedings of the international conference on computer vision. ICCV01. Vancouver, Canada, pp.763-770.
400. [Torralba03] Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America*, 20.
401. [Torralba03b] Torralba, A. and Oliva, A. (2003) Statistics of natural image categories. *Network* 14, 391-412
402. [Torralba99] Torralba, A.B., Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. In *International Conference on Computer Vision (ICCV99)*, pages 1253--1258, 1999.
403. [Treisman98] Treisman, A. (1998) The perception of features and objects. In R.D. Wright (Ed.), *Visual Attention* (pp.26-54). New-york Oxford : Oxford University Press
404. [Triesch03] Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, 3(1), 86-94.
405. [Tsotsos90] Tsotsos, J.K. (1990) Analyzing vision at the complexity level. *Behav. Brain Sci.* 13, 423-445
406. [Turano03] Turano, K. A., Gerguschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, 43, 333-346
407. [Tversky83] Tversky, B. & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology*, 15, 121-149.
408. [Unema05] Unema, P., Pannasch, S., Joos, M. & Velichkovsky, B.M. (2005). Time-course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473-494.
409. [Urbany89] Urbany, J.E., Dickson, P.R., Wilkie, W.L. (1989), "Buyer uncertainty and information search", *Journal of Consumer Research*, Vol. 16, September, pp. 208-15
410. [Vallacher99] Vallacher, R.R., Nowak A. (1999). The dynamics of self-regulation. In R.S. Wyer, JR (Ed.). *Perspectives on behavioral self-regulation* (pp.240-259). Mahwah, NJ : Lawrence Erlbaum Associates.
411. [Van Diepen98] Van Diepen, P. M. J., Wampers, M. and d'Ydewalle, G. (1998). Functional division of the visual field: moving masks and moving windows. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception*. Chapter 15, pp. 337-355. Amsterdam: Elsevier.)
412. [Van Trijp96] Van Trijp, H.M., Hoyer, W.D., Inman, J.J. (1996) Why Switch? Product Category: Level Explanations for True Variety-Seeking Behavior. *Journal of Marketing Research*, Vol. 33, No. 3 (Aug., 1996), pp. 281-292 *Journal of Marketing Research*, Vol. 33, No. 3 (Aug., 1996), pp. 281-292
413. [VanderHelm00] van der Helm, P.A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, 126, 770-800.
414. [Velichkovsky05] Velichkovsky, B.M., Joos, M., Helmert, J.R., & Pannasch, S. (2005). Two visual systems and their eye movements: evidence from static and dynamic scene perception. *CogSci 2005: Proceedings of the XXVII Conference of the Cognitive Science Society*. July 21-23 Stresa, Italy, pp. 2283-2288.
415. [Vertegaal99] Vertegaal, R. (1999). The gaze Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. *Proc. CHI1999, ACM*, 294-301
416. [Veryzer98] Veryzer, R. W., & Hutchinson, W. (1998). The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of Consumer Research*, 24, 374-393.
417. [Vincent07] Vincent, B., Correani, Baddeley, Troscianko, Leonards, (2007). Do we look at lights? Using mixture modeling to determine the contribution of low- versus high-level factors in visual scene exploration.
418. [Vogel06] Vogel, Julia., Schwaninger, Adrian., Wallraven, Christian, Bülhoff, Henrich, H. (2006). Categorization of natural scenes: local vs. global information. in *Symposium on Applied Perception in Graphics and Visualization APGV'06*, Boston, MA, USA. July 28-30, 2006.
419. [Wade05] Wade, N. J. & Tatler, B. W. (2005) *The moving tablet of the eye: The origins of modern eye movement research*. Oxford University Press.
420. [Walker77] Walker-Smith, G. J., Gale, A. G., & Findlay J.M. (1977). Eye movements strategies involved in face perception. *Perception*, 6, 313-326.

421. [Wandell95] Wandell, B.A. (1995). *Foundations of Vision*. Sunderland, MA: Sinauer. Palmer, S.E. (1999). *Vision Science Photons to Phenomenology*. Cambridge, MA: MIT Press
422. [Wang97] Wang, M.-J. J., Lin, S.-C., & Drury, C. G. (1997). Training for strategy in visual search. *Industrial Ergonomics*, 20, 101-108
423. [Watanabe03] Watanabe, K. (2003). Differential effect of distractor timing on localizing versus identifying visual changes. *Cognition*, 88(2), 243-257.
424. [Whitfield79] Whitfield, T. W. A., Slatter, P. E. (1979). The effects of categorization and prototypicality on aesthetic choice in a furniture selection task. *British Journal of Psychology*, 70, 65-75.
425. [Whittlesea90] Whittlesea, B. W. A., Jacoby, L. L., Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, 29, 716-732.
426. [Williams01] Williams, D. (2001) *Weighing the Odds: a Course in Probability and Statistics*, Cambridge University Press, 2001
427. [Williams98] Williams, P., Tarr, M.J. (1998). RSVP: Experimental control software for MacOS
428. [Winkielman00] Winkielman, P., Fazendeiro, T. (2000). Parallel effects of conceptual fluency on memory illusions and liking.
429. [Winkielman00b] Winkielman, P., Fazendeiro, T. (2000). The role of conceptual fluency in preference and memory. In preparation
430. [Winkielman01] Winkielman, P., Cacioppo, J.T. (2001). Mind at ease puts a smile on the face : psychological evidence that processing facilitation increase positive affect. *Journal of personality and social psychology*, 81, 989-1000
431. [Winkielman03] Winkielman, P., Schwarz, N., Fazendeiro, T., Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*. (pp. 189-217). Mahwah, NJ: Lawrence Erlbaum
432. [Winston92] Winston, Andrew S., & and Gerald C. Cupchik (1992), "The Evaluation of High Art and Popular Art By Naive and Experienced Viewers," *Visual Arts Research*, 18(Spring), 1-14.
433. [Witherspoon85] Witherspoon, D., Allan, L. G. (1985). The effect of a prior presentation on temporal judgements in a perceptual identification task. *Memory & Cognition*, 13, 101-111.
434. [Wolfe04] Wolfe, J. M. & Horowitz, T.S. (2004). What attributes guides the deployment of visual attention and how do they do it ? *Nature review neurosciences*, 5, 1-7
435. [Wolfe94] Wolfe, J.M., (1994). Guided search 2.0 : a revised model of visual search. *Psychonomic Bulletin & review*, 1, 202-238
436. [Wolfe96] Wolfe, J. M. (1996). Post-attentive vision. *Investigative Ophthalmology & Visual Science*, 37(3), 981.
437. [Wolfe98] Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology*, 8, R303-R304.
438. [Wooding02] Wooding, D. S. (2002). Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments and Computers*, 34(4), 518-528.
439. [Yantis 02] Yantis, S. (Vol. Ed.), Pashler, H. (Series Ed.) (2002). *Stevens' Handbook of Experimental Psychology*, 3e. Volume 1: Sensation and Perception. New York: Wiley.
440. [Yantis93] Yantis, S. (1993). Stimulus-Driven Attentional Capture and Attentional Control Settings. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3), 676-681
441. [Yarbus67] Yarbus (1967), *Eye movements during perception of complex objects*, in L. A. Riggs, ed., 'Eye Movements and Vision', Plenum Press, New York, chapter VII, pp. 171-196
442. [Zaichokwsky85] Zaichokwsky, J.L. (1985). Familiarity: Product use, involvement or expertise?. *Advances in consumer research*, 8, ed. Elisabeth C. Hirshman et Morris B. Holbrook, Association for consumer research, 296-299.
443. [Zajonc68] Zajonc, R. B. (1968) Attitudinal Effects of Mere Exposure. *Journal of Personality and Social Psychology*, 9, 2, 1-27.
444. [Zajonc72] Zajonc, R. B., Shaver, P., Tavis, C., & Van Kreveld, D. (1972). Exposure, satiation, and stimulus discriminability. *Journal of Personality and Social Psychology*, 21(3), 270-280.
445. [Zhang04] Zhang, Q., Imamiya, A., Go, K. and Gao, X. (2004). Overriding errors in speech and gaze multimodal architecture. *Proc. 9th International Conference on Intelligent User Interfaces (2004)*, 346-348
446. [Zinchenko62] Zinchenko, V.P., Chzhi-Tsin, V. & Tarakanov, V.V. (1962). The formation and development of perceptual activity. *Soviet Psychology and Psychiatry*, 3, 3-12.